FILE COPY

2

AD-A209 141

DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

DTIC
ELECTE
MAY 16 1989
S H D

1a. REPORT SECURITY CLASSIFICATION: Unclassified

1b. RESTRICTIVE MARKINGS: N. A.

2a. SECURITY CLASSIFICATION AUTHORITY: N. A.

3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE: N. A.

4. PERFORMING ORGANIZATION REPORT NUMBER(S): N. A.

5. MONITORING ORGANIZATION REPORT NUMBER(S): Same

6a. NAME OF PERFORMING ORGANIZATION: NAS-National Research Council

6b. OFFICE SYMBOL *(If applicable)*: CHABA

7a. NAME OF MONITORING ORGANIZATION: Office of Naval Research

6c. ADDRESS *(City, State, and ZIP Code)*:
2101 Constitution Avenue, N.W.
Washington, DC 20418

7b. ADDRESS *(City, State, and ZIP Code)*:
800 N. Quincy Street
Arlington, VA 22217-5000

8a. NAME OF FUNDING/SPONSORING ORGANIZATION: Office of Naval Research

8b. OFFICE SYMBOL *(If applicable)*: Code 1142PS

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-87-C-0272

8c. ADDRESS *(City, State, and ZIP Code)*:
800 N. Quincy Street
Arlington, VA 22217-5000

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61153N 42	RR 04209	RR 0420901	4424204

11. TITLE *(Include Security Classification)*: Classification of Complex Nonspeech Sounds (U)

12. PERSONAL AUTHOR(S): Yost, William; Braida, Louis; Hartmann, William; Kidd, Gerald; Kruskal, Joseph; Pastore, Richard; Sorkin, Robert; Warren, Richard

13a. TYPE OF REPORT: Final Report

13b. TIME COVERED: FROM 87/04/01 TO 88/08/31

14. DATE OF REPORT *(Year, Month, Day)*: 1989 April 14

15. PAGE COUNT: 96

16. SUPPLEMENTARY NOTATION: CHABA Panel on Classification of Complex Nonspeech Sounds

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*: Perception; Auditory Classification; Signal Processing; Auditory Object Perception; Limits of Auditory Processing Acoustic Transients; Sonar; Multidimensional Analysis.

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

The report reviews and evaluates all of the small amount of literature on the classification of complex nonspeech, nonmusic, transient sounds. Some of the literature on perception, signal processing, auditory object perception, limits of auditory processing, acoustic transients, sonar, and multidimensional analysis is also included to provide a meaningful matrix within which to imbed the subject report. Where possible the missing data, and research needs, are outlined in a general, unstructured fashion in order to encourage innovation in this important but infantile area. Are psycoacousticians finally going to leave the familiar security of pure tones, music, and speech in order to conduct research concerning the real world of meaningful, complex sounds?

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL: John J. O'Hare

22b. TELEPHONE *(Include Area Code)*: (202) 696-4502

22c. OFFICE SYMBOL: Code 1142PS

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE

Classification of Complex Nonspeech Sounds

Panel on Classification of Complex Nonspeech Sounds

Committee on Hearing, Bioacoustics, and Biomechanics
Commission on Behavioral and Social Sciences and Education
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1989

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Frank Press is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organisation of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Robert M. White is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Samuel O. Thier is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Frank Press and Dr. Robert M. White are chairman and vice chairman, respectively, of the National Research Council.

This work relates to Department of the Navy Contract N00014-87-C-0272 issued by the Office of Naval Research under Contract Authority NR 201-124. However, the content doest not necessarily reflect the position or the policy of the Department of the Navy or the government, and no official endorsement should be inferred.

The United States government has at least a royalty-free, nonexclusive and irrevocable license throughout the world for government purposes to publish, translate, reproduce, deliver, perform, dispose of, and to authorise others so to do, all or any portion of this work.

Available from

Committee on Hearing, Bioacoustics, and Biomechanics
National Research Council
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

Printed in the United States of America

PANEL ON CLASSIFICATION OF COMPLEX NONSPEECH SOUNDS

WILLIAM A. YOST (Chair), Parmly Hearing Institute, Loyola University
LOUIS D. BRAIDA, Research Laboratory of Electronics, Massachusetts Institute of Technology
WILLIAM W. HARTMANN, Physics Department, Michigan State University
GERALD D. KIDD, JR., Department of Communication Disorders, Boston University
JOSEPH B. KRUSKAL, Department of Mathematics, Bell Laboratories, Murray Hill, N.J.
RICHARD E. PASTORE, Department of Psychology, State University of New York, Binghamton
MURRAY B. SACHS, Department of Biomedical Engineering, Johns Hopkins University
ROBERT D. SORKIN, Department of Psychology, University of Florida
RICHARD M. WARREN, Department of Psychology, University of Wisconsin

DTIC COPY INSPECTED 4

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

COMMITTEE ON HEARING, BIOACOUSTICS, AND BIOMECHANICS

JOHN H. MILLS (Chair), Department of Otolaryngology, Medical University of South Carolina
ROBYN COX, Speech and Hearing Center, Memphis State University
ROBERT DOBIE, Department of Otolaryngology, University of Washington
C. DANIEL GEISLER, Department of Neurophysiology, University of Wisconsin
IRA J. HIRSH, Central Institute for the Deaf, St. Louis
WILLIAM M. RABINOWITZ, Research Laboratory of Electronics, Massachusetts Institute of Technology
JAMES C. SAUNDERS, Department of Otorhinolaryngology and Human Communications, University of Pennsylvania
ROBERT D. SORKIN, Department of Psychology, University of Florida
CONSTANTINE E. TRAHIOTIS, Department of Otolaryngology, University of Connecticut Health Center
DIANNE J. VAN TASELL, Department of Communication Disorders, University of Minnesota
HENNING E. VON GIERKE, Yellow Springs, Ohio

MILTON A. WHITCOMB, Study Director
ARLYSS K. WIGGINS, Administrative Secretary

Contents

Preface

In 1987 the Office of Naval Research (ONR) asked the Committee on Hearing, Bioacoustics, and Biomechanics (CHABA) to review and evaluate the literature on complex nonspeech sound processing by the human auditory system. CHABA established the Panel on Classification of Complex Nonspeech Sounds to review the literature and make recommendations for future research.

The primary focus of the panel's charge was a review and evaluation of the literature on labeling of very brief or transient sounds—a literature that turns out to be very small. The vast literature on detection, discrimination, and identification of sounds was not, however, reviewed. Rather than produce a report evaluating only a small literature on the labeling of brief sounds, the panel decided to include evaluation of related literatures it considered to be important to understanding the literature on labeling of brief sounds. Thus, this report includes a literature review of object perception (the term was chosen because it is more neutral than *streaming*, *figure/ground*, or *event perception*) and limits; in the panel's judgment these two areas contain important aspects of the overall task of labeling transient sounds. The panel also decided to review some informative literatures on nontransient sounds, such as music and speech, and to include some tasks other than labeling.

The report does not provide specific recommendations for future research. The panel considers the field of focus—the labeling of transient sounds—to be in its infancy and therefore believes that specific and highly structured research designs might unnecessarily limit innovation at this time.

As a review of a large section of the literature on the perception of complex sounds and an overview to aid funding in this area, the report should be of value to ONR and to others who are interested in the subject matter. In preparing the report, the panel has assumed that the reader is well acquainted with the area of hearing.

This report is the product of efforts by the entire panel. The summary of recommendations and the overview of research were drafted by the panel chair, with the advice and consultation of the panel members. The chapters that constitute the full review of the literature were drafted, section by section, by individual panel members who are expert in the particular fields; the depth of coverage in these sections therefore reflects their views on the relative importance of the topics as they pertain to the classification of complex sounds. Although the review of the literature contained in the report is not exhaustive, panel members attempted to review the major studies in each field.

The report is organized into six chapters: Chapter 1 is a summary of the panel's

recommendations for research on the classification of complex sounds. Chapter 2 is an overview and summary of the literature review, which is presented in greater detail in Chapters 3 through 6. Chapter 3 is a review of the limited literature on the classification of complex sounds itself. Chapter 4 is a review of research on auditory object perception. Chapter 5 deals with research on the limits of the auditory processing of complex sounds. Chapter 6 is a review of some of the research on speech perception.

In Chapter 3, the sections on the study of classification and multidimensional analysis were drafted by Joseph Kruskal; the section on classification of nonspeech transient sounds was drafted by Louis Braida; and the section on sonar detection by human observers was drafted by Robert Sorkin. In Chapter 4, the section on object perception and identification was drafted by William Hartmann; in the section on separating objects, the subsections on spectral profiles and spatial separation were drafted by William Yost and the subsections on temporal modulation and onset/offset characteristics were drafted by William Hartmann; the section on perception of temporal patters was drafted by Richard Warren; and the section on general principles of perceptual organization was drafted by William Yost. In Chapter 5, the section on the role of memory was drafted by Louis Braida; the section on uncertainty and attention was drafted by Gerald Kidd; and the section on limitations due to internal noise was drafted by Robert Sorkin. In the section on learning, the introduction and subsection on learning of complex nonspeech and nonmusic sounds were drafted by Gerald Kidd; the subsections on psychophysical abilities, discrimination of tone sequences, categorization of speech and music, second language acquisition, musical illusions, and perceptual learning were drafted by Richard Pastore; and the subsection on Morse code learning was drafted by Joseph Kruskal. Chapter 6, on speech perception, was drafted by Richard Pastore.

Although the work of drafting specific sections was divided among panel members, responsibility for the report is shared by all. I want to thank the members of the panel for their time, their expert knowledge, and their cooperation in this effort.

William A. Yost, Chair
Panel on Classification
of Complex Nonspeech Sounds

1
Recommendations for Future Research

This report provides a review of the basic literature on perception and classification of complex sounds and makes general recommendations for future research, especially regarding issues of classification.

DEFINING CLASSIFICATION OF COMPLEX SOUNDS

Because the subject of this report is classification of sound in general, it does not cover in depth classification of special sounds, such as those of speech and music. Although examples from these two areas are abundant in the report, our general concern is with nonspeech and nonmusical sounds in order to survey classification of all sound.

Considerations of human interaction with sound and a review of the existing literature suggested to the panel that most meaningful sounds of every day life have the following general properties:

(1) *Spectral complexity*: The sounds almost always consist of more than one frequency component.

(2) *Temporal complexity*: The sounds are time-varying; therefore, the spectral and temporal characteristics of the sound vary over the duration of the sound. Sometimes the change in a sound marks the beginning or end of an important aspect of the sound.

(3) *Brevity*: The information-bearing elements of most sounds last less than a second. Even in cases in which a sound is longer or is part of a sequence of sounds, the attribute of the sound to be classified lasts for a brief period of time.

(4) *Sound embedded in noise*: The sound of interest is often embedded in an acoustic background that contains the sounds of many sources.

These four criteria form the panel's definition of a complex sound. Although, in reviewing the literature we attempted to focus on studies involving complex sounds (as defined above), much of the literature on simple sounds was also reviewed because of its relevance to the overall problem of the classification of complex sounds.

In describing the perception of sounds the terms *detection*, *discrimination*, *identification*, *recognition*, *categorization*, and *classification* are used. It is not possible to provide exact definitions of these terms that are adhered to uniformly throughout the literature reviewed. The general definitions, used by the working group to formulate its recommendations, refer to *detection, discrimination, identification*, and *recognition* as processes in which listeners select particular sounds from a defined set of sounds (usually a small set of

candidate sounds) without the requirement that they label the sounds, while *catergorization* and *classification* refer to labeling sounds directly (often from a large set of sounds). Labeling refers to placing the sounds into different categories or directly assign names to sounds. These labels or categories can refer to a physical property of the sound (e.g., frequency), a subjective attribute of the sound (e.g., timbre), the source of the sound (e.g., a hand clap), a use or effect of the sound (e.g., unpleasant), or a special label (e.g., phoneme, musical chord name). Obviously most sounds can be categorized or classified with a variety of such labels.

With regard to reviewing the literature on the classification of complex sounds, the panel defined its task as: to provide a review of the literature on the labeling of sounds that meet most of the four criteria for the sounds of everyday experience (sounds that are complex, brief, time-varying, and embedded in a complex acoustic background). If the literature on music and speech is excluded, only a small body of literature remains that concerns the classification of complex sounds.

In the absence of much literature on the classification of complex sounds, the panel examined more closely the task confronting a human who attempts to classify (label) a complex sound. This classification task requires that the listener perceives the sound to be labeled as distinct from other sounds. The panel refers to the perception of a particular aspect of a complex sound as *auditory object perception*. Many, if not most, sounds that are to be classified form auditory objects. As an example, the noise from a busy street may contain the sound of a braking car. The sound of the braking car may become perceptually isolated as an auditory object. Although the auditory object may be labeled (e.g., as a high-frequency squeal, or as a braking car, or as unpleasant, or as dangerous), the formation of auditory objects does not require that listeners attach a label to the perceived sound.

A listener to speech or music is often called on to recognize a sound, that is, to state which of a limited ensemble of possible sounds is represented by the perceived acoustic message. There are many sounds in the natural environment, however, for which we would like listeners to characterize some attribute not so easily specified by an ensemble of possible messages. The response is more open and the listener may be asked to use adjectives of his or her own choosing that may refer to physical attributes of the sound, identification of a sound source, or description of some set of qualities of the sound. These processes are what is meant by classification. Similar procedures have been used to describe the stimulus space or subjective space of sounds in other modalities, like the dimensions of taste, smell, pain, etc. Within a complex sound, most often when it is presented as part of a stream or in competition with other background sounds, the listener is often asked to attend to or to report on a particular sound, a set of sound features, or perhaps a recognitionresponse that indicates what the sound sounds like. Such a task leads us to define an auditory object.

Auditory object perception in a complex sound field is a major component of complex sound classification. There is a substantial literature, which the panel attempted to review, on the topic of auditory object perception. Although some of that literature concerns studies in which listeners label sounds, in many of the articles reviewed by the panel listeners were not directly asked to label auditory objects.

The nervous system obviously imposes limits on the ability to perform any auditory task. The sensitivity and the resolving capabilities of the auditory system describe the limits of the system's processing power. Auditory processing is also limited by memory, learning, uncertainty, attention, internal noise, etc. Knowledge of these limits is essential for understanding the classification of complex sounds. Therefore, this report reviews the literature concerning these limits and their application to complex sounds.

Although the panel did not set out to review the literature on speech perception, we

found that all the issues discussed above have been covered in this literature. A review on speech perception is included for at least two reasons: (1) it provides a review of the speech literature that is germane to the issues described above and (2) the use of a particular sound (speech) provides a means of illustrating many of the points made in the other sections of the review.

In the absence of a large literature on understanding the entire process of classifying complex sounds, the par .l focused on auditory object perception as a major ingredient in classifying sound. And we felt that understanding the process of classification of complex sounds required an understanding of the limits imposed by the nervous system on auditory processing.

RECOMMENDATIONS FOR FUTURE RESEARCH

The study of auditory processing of complex sounds, especially those encountered in the real world, has become an increasingly significant area for auditory investigation. A great deal of the previous work in hearing dealt with simple sounds (e.g. sinusoids, clicks, noise bursts), most of which are not found in the everyday world. Classification of complex sounds is an important aspect of this newer area of investigation. Understanding the formation of auditory objects and the limits of processing complex sounds is crucial in studying the classification of complex sounds. Experimental and theoretical studies of these topics will not only provide valuable insights about hearing, but they will also bring much of the basic knowledge closer to practical application. As auditory science attempts to understand the processing of real-world sounds, many contributions will be made that have a direct practical impact on society. The panel therefore recommends that a significant effort be made to support research on the topic of classification of complex sounds, as defined in this report. The rest of this report highlights some topics that appear to be particularly crucial for study at this time. The remaining chapters of this report should be consulted to clarify terms and provide the background for these recommendations.

The recommendations given below provide general guidelines rather than lists of suggested experiments. The panel feels that hearing science has just scratched the surface of the issues of the classification of complex sounds. It is, we feel, premature to suggest specific studies. General guidelines should serve as a vehicle for selecting particularly fertile research projects. In addition to providing some general guidelines, this section poses some questions that appear important to answer at this time, some justifications for supporting certain topics, and some caution concerning investigation of some areas. This approach should allow for the creative and unusual proposal to surface more readily. The field is ripe for innovative and perhaps unorthodox projects. The panel hopes that the recommendations point the direction toward fruitful questions to be investigated without unduly limiting inquiry into how humans classify complex sounds.

Auditory Object Perception

As the literature review reveals, finding auditory objects in a complex sound is a major aspect of hearing. Little is known about how the auditory system accomplishes this task. Research into the role of intensity profiles, harmonic structure, onsets and offsets, temporal patterns, and spatial separation, as variables which help parse a complex sound into its probable sources, appears to promise significant advances in knowledge about each of these suggested ways to form auditory objects. Research into understanding the interaction of

these (and other) means of forming auditory objects is also required. For instance, are some cues used to form auditory objects more salient than others?

One way to demonstrate that a particular manipulation generates an auditory object is to show that listeners can perceive the object under simulated conditions. For instance, can a complex sound be presented over headphones so that various sound sources are spatially separated and external to the listener, as they are when a person listens to the sound sources in a real sound space?

The study of object perception appears particularly likely to yield practical applications. A great deal of effort is being put forth on machine and human-machine recognition of complex sounds. Devices, from speech recognizers to sonar detectors, are being developed to mimic or perhaps improve the human ability to process sound. Humans are often much better than machines at processing complex sounds. A major task for many of these devices is to find a particular type of sound or sound source (e.g., a sonar echo representing a ship) in a complex sound environment. Gaining a better understanding of how the auditory system forms auditory objects may provide ways to improve existing devices or suggest new ones.

Research on object perception at both the basic and applied levels can benefit from knowledge gained from studies of music and speech perception. An approach sometimes used in the study of speech and music perception may prove useful for other complex sounds: in this approach, both the physical properties of the sounds and the responses of the listeners are subjected to some form of multidimensional analysis. Combining the analysis of the stimulus with that of the response can facilitate the discovery of orderly, but complex, relationships between the stimulus and the response. Once these relationships have been identified, the physical properties of the sounds can be altered along the lines suggested by the physical analysis and the responses can be obtained again to test if the predicted changes in the responses occur. This method may provide a framework to assist in applying a relationship identified in one experimental context to those that may be discovered in other contexts.

The study of temporal sequences, as one aspect of auditory object formation, has generated a great deal of data. However, these data need to be integrated into some form of a quantitative account or theory. For instance, is there a way to account for the large range of durational limits for processing sequences and identifying temporal order, or can the phenomena of restoration and streaming be integrated into one theory?

Limits of Auditory Processing

As a general recommendation for future research, more must be learned about how memory, attention, uncertainty, learning, and internal noise limit auditory processing of complex sounds. The first logical step is to build on past work involving simple stimuli and on research conducted in other areas of science, such as vision. Some of this past work shows that a prior knowledge on the part of the listener is an important variable in classifying sounds. This suggests that both "top-down" and "bottom-up" modes of processing are important in tasks involving complex sounds. Little is known about the possible hierarchical approaches to processing complex sounds. Do variables such as the range of stimuli that the listener must process, or the multidimensional complexity of the stimuli, or the physical aspects of certain stimuli (e.g., those near the edges of the range) determine the nature of the limits imposed by memory or uncertainty? Answers to these questions will provide valuable insights into the classification of complex sounds and might also help unify theories of memory and uncertainty across the senses.

As stimulus complexity increases, uncertainty concerning the dimensions of this complexity has been shown to limit performance. But how does uncertainty limit auditory processing and classification, to what degree is classification limited by uncertainty, and in what way can these limits be overcome? For instance, how and to what degree can specialized training overcome limits imposed by uncertainty?

The study of learning in tasks involving complex sounds could lead to a number of payoffs. Such knowledge could improve our understanding of learning in general; it would clarify the extent to which learning is a limit for classifying complex sounds; it would provide a basis for training programs involving learning novel sounds; and it would help determine the extent to which learning is a major component in speech and music perception. Support of learning research should include studies of the effects of long-term learning or experience. A listener may be able to obtain a certain performance level, but only after prolonged practice. If a great deal of learning is required to master certain tasks, then under what conditions does learning take place, and is there a way to shorten the learning time?

Methodologies and Theories

Many of the excellent methods and theories currently used in auditory science were developed to describe or predict auditory processing of simple sounds, and therefore they may have limited application for studying complex sounds. One current approach, the detection-recognition theory and procedure, appears promising as a way to form a bridge between studies of simple stimuli and those that might be conducted using complex sounds.

The multidimensional nature of complex sounds coupled with a large response repertory demands either that the more established methods be expanded or that new methods be developed. Among the new developments some are quite likely to stem from studies of multidimensional scaling (MDS).

The development of new methods for studying complex sound processing may involve adapting techniques from other areas. The scientific study of classification in other scholarly fields (i.e., the development of classification schemes), procedures used in the study of visual perception, and methods used in music and speech research are three areas cited in the literature review that may provide valuable new insights for the study of the classification of complex sound. Methods for evaluating performance in the classification of complex sounds would provide a valuable research tool and might also be useful in practical situations requiring evaluation of major human-machine systems.

Support for methodologies or theories per se is often difficult to obtain. However, in the case of complex sounds, development of new methods, theories of data analysis or interpretation, or measurement tools would provide a valuable contribution. Work in this area should not be limited to theories of a particular phenomenon, although these too are needed, especially in the areas of temporal sequences, memory, and object perception. For instance: Are there classes or types of theories or accounts that have been applied in other areas of science that can be applied to the study of complex sounds? Are there methods for combining theories into one structure that would facilitate integration of accounts of seemingly disparate phenomena?

Other Areas for Future Research

The literature review documented many studies indicating large individual differences in the performance of some tasks involving complex sounds. The causes of individual differences and the correlations among tasks will be important knowledge for understanding

human classification of complex sounds. For practical situations, understanding the nature of individual differences may assist in designing screening procedures to select individuals with certain performance abilities or to determine those individuals who are deficient in some ability.

Studies that directly investigate labeling complex sounds may provide valuable results. However, the literature review indicates that some studies of a fixed set of sounds have not led to results that can be generalized to other conditions. Direct investigation of the labeling of specific complex sounds will prove most useful if they are guided by theory so that the results can be applied to a wide set of conditions.

This report is focused on auditory perception; however, it is clear that more must also be learned about the central auditory nervous system. The relationship between perception and the structure and function of the nervous system will have to be clarified before any complete theory of auditory classification is possible. The study of animal models will most likely play an important role in linking our knowledge of perception to that of the nervous system. As support is being supplied for understanding the microbiology of the various parts of the auditory system, so should significant support be given for establishing connections between perception and neural structure and function.

2
Overview and Summary of the Literature

In this chapter we highlight the major points made in the literature reviewed, which is covered in detail in Chapters 3–6. This review forms the background for the panel's recommendations for future research, described in Chapter 1.

The first section of this chapter (which corresponds to Chapter 3) covers the limited literature on the classification of complex sounds. It begins with a review of work done on the general topic of classification as used outside the fields of perception. Many scholars study the topic of classification, which is often used in a different way than it was defined in Chapter 1. The rest of the section focuses on some work involved with the classification of nonspeech transient sounds and sonar detection. These topics represent the literature that appeared most germane to the entire process of the classification of complex sounds as described in Chapter 1.

The next section of this chapter (which corresponds to Chapter 4) covers the topic of auditory object perception. A number of terms have been used almost synonymously with object perception: *entity perception, source perception, event perception*, etc. In the context of this report, auditory object perception refers to those processes that allow one sound to be separated from other sounds. As such the terms *streaming* and *stream segregation* (see Bregman, 1978a) are also seen as approximate synonyms for object perception.

The next section of this chapter (which corresponds to Chapter 5) covers the limits of auditory processing. The particular limits that the panel feels are crucial for understanding classification of complex sound include memory, uncertainty, attention, internal noise, and learning.

The final section of this chapter (which corresponds to Chapter 6) summarizes a review of some of the speech perception literature, specifically topics in the speech literature that appeared most relevant to the general issue of complex sound classification as described in Chapter 1.

CLASSIFICATION OF COMPLEX SOUNDS

Both the general study of classification (as used in fields other than perception) and some work on the classification of complex sounds are reviewed. The work on classification of complex sounds provides examples of some of the limited attempts that have been made at studying complex sound classification. Other examples are reviewed in Chapters 4 and 5

because these studies deal with auditory object perception or the limits of complex sound processing.

The Study of Classification

In recent decades, the general topic of classification has received a great deal of systematic study in a number of different disciplines. This is evident in the many books on the subject, the formation of approximately eight scientific societies for which classification is the prime topic, the creation of the *Journal of Classification*, and the recent formation of the International Federation of Classification Societies.

In these fields, the term *classification* refers to creating a classification, also commonly referred to as a clustering or a taxonomy. It does not refer to the closely related task of deciding to which class (in a preexisting classification) an entity belongs. The latter problem is often referred to as classification in statistics or categorization in the study of speech, while the word *discrimination* (which has a different meaning in the perceptual literature) is preferred in the classification literature. In the classification literature, a classification is taken to be either a simple classification, (i.e., a partition into mutually exclusive and exhaustive classes) or a hierarchical classification (e.g., a biological taxonomy), although numerous other variations have received attention.

Multidimensional Analysis

One technique sometimes used to study complex sound classification is multidimensional scaling (see Carrol and Kruskal, 1978). Multidimensional analysis, in connection with the kinds of data considered in auditory research, consists of two kinds of models and several techniques for representing complex sounds. One kind of model is spatial: a low-dimensional (often Euclidean) space serves as the model, and each sound is represented by a point in the space, such that the distances between points correspond to perceptual similarities between sounds. The other kind of model is set-theoretic: a set of abstract features serves as the model, and each sound is represented by the set of features it possesses.

Although neither type of model is valid as a complete description of how people function, both models are capable of providing useful insights into such function. There is also no conflict between the two models. In some cases both can be used to good advantage on the same data, and they often provide different sorts of information. Thus, the use of each type of model should be based on its strengths and weaknesses as applied to each situation.

Classification of Nonspeech Transient Sounds

Howard and his associates (Howard and O'Hare, 1984) have undertaken studies focused directly on the classification of nonspeech transient sounds. These studies are characteristic of those that have directly investigated the classification of complex sounds. The sounds studied ranged from actual sounds recorded underwater (as germane to submarine sonar detection) to temporally and spectrally shaped noises that were intended to mimic the sounds of actual sources.

Two sets of investigations are particularly relevant. In one series of studies, real-world and synthetic sounds were analyzed (using a simple auditory model) to determine which physical properties formed the basis of similarity judgments. In general, relatively simple properties, as might be conveyed by low-order principal components of spectral shape such as tilt and compactness, correlate with the more complex dimensions revealed by

multidimensional scaling analysis of similarity ratings. Generally, listeners with musical training were more influenced by temporal properties of the sounds (periodicity and pitch) than by spectral shape properties, in contrast to listeners without such training.

The second series of studies focuses on the role of syntactic (structural) and semantic (interpretive) factors indetermining the ability of listeners to distinguish specified sound patterns from randomly constructed patterns. Typically, listeners appear to utilize the syntactic structure provided by a simple finite-state grammar to improve the rate at which they learn to discriminate sound sequences and/or to improve the final level of performance they can achieve. The effect of semantic themes is more problematic. For some listeners, instruction that provides thematic interpretation of sound patterns improves the discriminability of grammatical sequences; for others, it merely facilitates the initial learning of the discrimination task. The task of classifying sonar returns has received some attention in the unclassified, nonmilitary literature (e.g., Howard and Silverman, 1976; Kobus et al., 1986). A number of studies have attempted to define how the classification of these complex nonspeech sounds depends on the physical properties of the signals and on the listener's training, knowledge, and expectations. In general, these studies have not revealed any fundamental new information that differs from that observed in experiments on auditory perception of simple sounds and speech perception.

AUDITORY OBJECT PERCEPTION

The text by Moore (1982) provides a useful introduction to the topic of auditory objects and patterns. Moore organizes the subject in terms of: (1) object perception and identification, (2) separating objects, (3) perception of temporal patterns, and (4) general principles of perceptual organization. The literature review in this report is organized in a similar manner, but it includes research in addition to that considered by Moore. McAdams (1984a) and Hartmann (1987) also provide useful reviews of the auditory object perception literature.

Object Perception and Identification

For sounds consisting of a single frequency, three parameters are necessary for identification: frequency (pitch), intensity (loudness), and duration. Those humans who do not have perfect (absolute) pitch can identify only 5–6 simple sounds out of a large set of tones varying in frequency, intensity, or duration. As a sound's spectral complexity (number of frequency components in the sound) increases, the number of possible classifications also increases. For complex sounds, spectral complexity, sometimes associated with the percept of timbre, is generally used to describe the sound. The spectrum, and thus the timbre, can be static or dynamic over time. In time-varying patterns, such as the sounds of musical instruments, onset characteristics are often crucial for identification. Temporal instabilities in the steady-state portion of sounds may also aid in object formation, but this variable has received little attention in the literature.

Separating Objects

In considering complex sounds, the sound field may consist of many sources. It is not clear how the spectral and temporal properties associated with each source are represented within the auditory system, since the entire sound field is presumably coded at an early stage of auditory processing. A number of acoustic properties have been suggested as

responsible for allowing the system to classify the various sound sources that may make up a complex sound. Some of those properties are spectral profile, temporal modulation, onset/offset disparities, and spatial separation.

Spectral Profile

The work of Green and his colleagues (see Green, 1988 for a review of this work) has demonstrated the importance of the contour of amplitudes in the spectrum of a complex sound for discriminating among different stimuli. Sounds with subtle changes in the amplitude profile can be discriminated despite large random variations in the overall amplitude of the sound.

The various models of complex, or virtual (Terhardt, Stoll, and Sweewann, 1982), pitch are based on various forms of spectral pattern recognition (see de Boer, 1974, for a review). The spacing of the components in a complex spectrum is a major determinant of the pitch and, to some extent, the timbre of the sound.

This empirical and theoretical work indicates that small differences in the amplitudes of spectral components and in the spacing of the components of a complex sound may be a basis for forming auditory objects.

Temporal Modulation

Many sound sources have a slow amplitude and frequency modulation of the primary vibration pattern. Consider the vocal cords: the pulse rate of the vibrating vocal cords determines the fundamental frequency, or pitch, of the voice. However, the period of the pulses is not perfectly constant, nor are the amplitudes of the pulses always the same. These changes produced in the vocal cord pulses result in a frequency and amplitude modulation of the fundamental frequency associated with the mean pulse rate of the cords. Most sound sources have such temporal modulations, and it is possible that the pattern of these modulations is unique to each sound source. A number of recent studies have shown that these forms of temporal modulation, are used to help classify sounds into their probable sources. For instance, different voices in a mixture of voices can be recognized as separate if a unique pattern of frequency modulation (vibrato) is added to the waveform for each voice (McAdams, 1984b).

If the auditory system is to process these slow temporal modulations, then it might operate as a wide band detector. As such, there might be significant interactions among frequency channels when the task involves temporal modulation processing. The work on comodulation masking release (CMR, see Hall, 1987) is an example of such an interaction. In CMR the detection of a masked signal in a narrow band of noise is improved if another narrow band of noise with the same temporal structure as the masking band is presented in a different frequency region than that of the masker. Yost and Sheft (1988) have shown that the ability to detect amplitude modulation of one tone can be significantly interfered with when another tone is also amplitude modulated.

Onsets and Offsets

The nature of the rising and falling parts of a sound can be the sole basis for their eventual classification. The best examples of this cue for object perception are from music synthesis. In synthesizing different instruments playing the same pitch, it is common to vary mainly the transient characteristics of the sound to simulate the particular instrument.

Auditory Space

Spatial separation of sound sources promotes the perceptual separation of auditory objects, as shown in the cocktail-party effect experiments by Cherry (1953). Dichotic pitch phenomena (and perhaps studies of the masking-level difference, Green and Yost, 1975) may be regarded as the separation of a tone from a noise background, on the basis of interaural differences as spatial cues (see Yost, Harder, and Dye, 1987, and Hartmann, 1987, for reviews). But spatial separation does not guarantee perceptual separation of the objects. Studies of speech, music, and pitch perception indicate that in some cases binaural disparities lead to a spatially fused percept, rather than to a perception of separation.

Perception of Temporal Patterns

There is an extensive literature on the temporal properties of sound that assist in the formation of auditory objects. The experiments in this area tend to fall into three categories: (1) stream segregation, (2) perception of sequential patterns, and (3) perceptual restoration of sequential sounds. These three groups are not always mutually exclusive.

Stream Segregation

The concept of streaming concerns the tendency for certain sequences of sounds in a complex sound field to appear as one object, as if this sequence were a stream isolated from other sounds (Bregman, 1978a). Sound sequences with corresponding spectral, spatial, intensive, and temporal characteristics often form such streams. Despite a number of articles on this topic, the requisite characteristics are often difficult to quantify, and no comprehensive theory has emerged for predicting when a sequence will form a stream.

Sequential Patterns

For sequences of sounds, listeners can be asked to discriminate among different arrangements of the same sound or to identify the components in a sequence (the identification task usually requires that the listeners also report the order of the items). A common goal of discrimination and identification is to measure the minimum duration required for the assigned task. This duration is then used to estimate the integration time of the auditory system for processing sequential information. The reported estimates of integration times range from a few milliseconds to several seconds, depending on the nature of the task, with component identification requiring longer item durations than discriminations involving permuted orders. However, there is no generally accepted temporal theory that consolidates these various estimates or relates them to other measures of the temporal integration period for auditory processing (see Green, 1971; Hirsh, 1976; Moore, 1982; and Warren, 1982, for reviews).

Restoration in Sequential Sounds

Disruption of a signal by a louder extraneous sound can lead to auditory induction (Warren, Obusek, and Ackroff, 1972) or pulsation thresholds (Houtgast, 1972). For these conditions, a signal interrupted by a louder sound may appear to be continuous. It is as if the auditory system restores the missing sound during the period when it is absent. Warren (1984) has classified restoration into three types: (1) heterophonic continuity (this category includes pulsation thresholds), which involves the illusory continuation of

one sound when interrupted by a different (e.g., a different frequency), louder sound; (2) homophonic continuity, which is the illusory continuity of a sound when interrupted by a louder level of the same sound; and (3) contextual concatenation, which does not involve illusory continuity of a steady-state signal (as do the other types of auditory inductions) but consists of restoration of an item that differs from the preceding and following sounds. An interesting variety of contextual concatenation is phonemic restoration, in which missing speech segments are restored in keeping with the application of syntactic and semantic rules. Again, no comprehensive theory has emerged for predicting when these forms of restoration will occur. Nor is there an adequate understanding of the relationship between perceptual restoration and other sequential phenomena, such as streaming.

General Principles of Perceptual Organization

The literature provides only a few hints of general principles for perceptual organization. The Gibsonian view argues that perceptual classification is based on knowledge about the sources that generate the sound as much as on the sound itself. The work of Bregman and his colleagues on stream segregation is largely an attempt to describe properties of sound that may form figure (foreground) and ground (background) in a complex sound field. Both the ecological approach of the Gibsonians and the hypotheses concerning the formation of auditory streams are founded in, or at least consistent with, Gestalt principals.

LIMITS OF THE AUDITORY PROCESSING OF COMPLEX SOUNDS

The auditory system is limited in its ability to process sounds by memory constraints, by learning, by uncertainty concerning the possible stimulus and response sets, and by various forms of internal noise. Although these limitations to auditory processing have received considerable attention in the auditory (and visual) literature, the focus of the work has been on simple sounds. Understanding the limits imposed on processing complex sound has received less attention.

Memory

The task of identifying or recognizing particular sounds from a set of sounds when the sounds are relatively simple (e.g., tones of different levels) has generated numerous experiments and a few theories. Many of the theories have come from vision and areas of verbal learning.

The performance of listeners in many sound identification tasks is determined to a large extent by the range over which the stimuli vary and by the characteristics of the stimuli at the edges of this range. Recent analytic models (e.g., Braida and Durlach, 1986) have synthesized a great deal of the data based on simple stimuli. These models are phrased in terms that can be extended to more complex stimulus conditions. For these complex conditions an important variable appears to be the number of stimulus dimensions that covary across the stimuli that are to be identified. The greater this covariance, the better able listeners are at identifying the stimuli.

Uncertainty and Attention

In general, uncertainty about the spectral or temporal structure of a sound interferes with the listener's ability to extract information from, or about, the sound and its source.

Early studies of uncertainty effects with simple sounds demonstrated small, but consistent, reductions in performance when some aspect of the sound or its presentation was uncertain.

Much larger reductions in performance are found when the stimulus becomes more complex. Watson and his colleagues (see Watson, 1987, for a review) have demonstrated large changes in performance when uncertainty is systematically varied in tasks involving 10-tone sequential patterns. The listener's task is to determine if one of the 10 tones in a comparison 10-tone pattern is different from that in a standard pattern. When the set from which these patterns is chosen is large and the patterns and components subject to change are randomly selected, then performance can be degraded by a large amount relative to cases involving small sets. Similar, but less severe effects of uncertainty have been obtained with the spectral profiles used in the studies by Green and his colleagues (see Green, 1988, for a review). Directing the observer's attention to the crucial element of a complex sound may reduce the effects of uncertainty (e.g., Watson, Kelly, and Wrotson, 1976; Howard et al., 1984).

Internal Noise

Many of the decrements in performance measured in the tasks cited above can be modeled by assuming that performance is degraded by the addition of an internal noise in the sound processing. Models of internal noise for detection, and to some extent, discrimination and identification of simple stimuli have been proposed for many years (for a review see Gilkey and Robinson, 1986). A paradigm that is often used is the "frozen noise" procedure, in which the same stimulus is presented on every trial. Variations in performance are assumed to be due to internal noise because there is no variability in the external stimulus. The internal noise can be introduced at the site of transduction, at the site of stimulus processing, or at some decision stage. Although internal noise models have been successful in accounting for data involving simple stimuli, less work has been done in predicting data using complex sounds.

Learning

By far the most frequent reference to or use of the term *learning* in the literature on nonspeech sound perception is in acquainting listeners with the requirements of a particular experimental task or with internalizing the value of a stimulus along a particular perceptual dimension to be used as a reference. In contrast, learning to attend to specific aspects of a complex sound or sound sequence that varies along several dimensions simultaneously, and attempting to assign the stimulus to a particular group, has not been studied extensively.

The issues involved in learning in audition are complex and diverse, extending across multidisciplinary boundaries. Learning of special sounds, such as music, sonar returns, speech, a second language, and Morse code, have been studied by a variety of different scientists. An interesting theme emerging from some of these studies is the notion of different processing strategies based on the temporal properties of the sound or sounds to be identified. The identification of steady-state sounds may involve more bottom-up processes because of the time available to extract critical stimulus features. Transient sounds, by comparison, cannot be analyzed in that manner and may depend to a greater degree on prior knowledge about the structure and the likely source of sound.

Recent research on learning nonspeech auditory patterns (Leek, and Watson, 1984) has revealed some important constraints governing listeners' abilities to learn such patterns. The amount of uncertainty in the stimulus and the way in which the various acoustic cues

are packaged within the stimulus sets are crucial elements in determining how many items a listener can learn.

LESSONS FROM SPEECH PERCEPTION

Speech is one class of acoustic stimulus for which classification (usually referred to as categorization in the speech literature) has been studied extensively for many decades. The relevance of the speech research literature to the study of categorizing nonspeech sounds depends on assumptions concerning the nature of speech perception. One assumption is that speech perception is based on some form of processes unique to human speech mechanisms. If this view is valid, then the extensive speech perception literature may provide examples only of strategies and techniques for the study of categorization. However, some researchers assume that many of the apparent perceptual differences between speech and other acoustic signals may be artifacts of the largely independent development of the research fields (e.g., Diehl, 1987; Pastore, 1981; Pisoni, 1987; Schouten, 1980). These researchers maintain that speech perception may be based on higher-order stimulus processing, which is largely learned and has developed, at least in part, to make use of unique properties of the human auditory system. If this latter view is valid, then much if not all of the extensive literature on speech perception may be directly relevant to the classification of complex sound.

Much of the human speech perception research has focused on the relationship of categories of perception to both the acoustic stimuli of speech and the structures of production (or articulation) that normally produce the acoustic stimuli. This study of the relationship among (a) the characteristics of the sound production source, (b) spectral and temporal properties of sound, and (c) categorical properties of perception represents a type of working structure for future studies of categorization of naturally produced acoustic stimuli (e.g., animal calls, engine noises, speech and speaker recognition), whereas the source properties probably are not important for the categorization of artificially coded cues (e.g., types of alarms, cues for the status of equipment, or even the recording of information by equipment monitoring aspects of the environment).

Even if the data obtained with speech stimuli are assumed to be of limited value for studying other complex sounds, many of the procedures used to study speech perception and some of the theories may provide valuable tools and insights. For instance, the research procedures used to study categorical perception and the models that address the findings from these procedures maybe applicable to the general issue of the classification of complex sounds.

3
The Classification of Complex Sounds

THE STUDY OF CLASSIFICATION

In recent decades, the topic of classification has received a great deal of systematic study. This is evident in the number of recent books on the topic, the formation of approximately eight specialized scientific societies, the creation of the *Journal of Classification* in 1984, and the recent formation of the International Federation of Classification Societies, which held its first meeting in Aachen, Germany, in 1987.

In this field, the term *classification* normally refers to creating a classification, also commonly referred to as a clustering or a taxonomy. It does not refer to the closely related problem of deciding to which class (in a preexisting classification) an entity belongs. Although the latter problem is sometimes referred to as classification in statistics and other fields, the term *discrimination* is preferred in the classification literature.

Generally, a classification is taken to be either a simple classification (i.e., a partition into mutually exclusive and exhaustive classes) or a hierarchical classification (e.g., a biological taxonomy), although numerous other variations have received attention. Often a hierarchical classification is created as one step toward a simple classification. The most common form of data used is a matrix of objects (individuals, entities, etc.) by variables (characters, etc). The second most common form of data used is a square (usually symmetric) matrix of proximities (similarities, dissimilarities, distances, etc.) among the objects. Sometimes data of the former type is converted into data of the latter type by some mathematical procedure as a preliminary operation and the latter used to create the classification. However, numerous other types of data have been considered.

In the early literature on classification, the most common topic was new methods for making classifications, and a great many methods were proposed in different fields. As people discovered each other's work, it became important to compare these methods and see how they related to each other. The purpose of making the classification was recognized as important: some classifications are used for administrative convenience (e.g., classification of city locations into police precincts), some are intended to improve performance (e.g., classification of red spotted diseases to improve treatment); some are intended to improve understanding (e.g., the Linnean taxonomy); and so on. Some classifications (e.g., into voting districts) may reasonably impose classes where they do not previously exist, while others (e.g., into species) are intended to reflect an underlying reality.

Today, the topic that engages greatest attention is the determination of the properties of methods and classifications. As an example, one question asked using data subsampling

(e.g., split halves, jackknife, bootstrap) is how stable a classification may be. Another, using mathematical analysis, is how the method would perform as the amount of available data became indefinitely large.

Existing societies devoted to classification include the Classification Society of North America, the Society for Numerical Taxonomy, the British Classification Society, Gesellschaft fuer Klassification, the Japan Classification Society, Societé Francophone de Classification, as well as organizations in Italy and Yugoslavia.

The earliest modern book on classification is Sokal and Sneath (1963), which played a major role in stimulating the modern surge of interest in the subject. The 1970s yielded Jardine and Sibson (1971); Blackith and Reyment (1971, although it is not directly on the topic); Sneath and Sokal (1973, a revision of the 1963 book); Anderberg (1973); Bock (1974, in German); and van Ryzin (1977, the proceedings of a conference).

MULTIDIMENSIONAL ANALYSIS

Multidimensional analysis, in connection with the kinds of data considered in this report, consists of two kinds of models and several techniques for representing complex sounds. One kind of model is spatial: a low-dimensional Euclidean space serves as the model, and each sound is represented by a point in the space. The other kind of model is set-theoretic: a set of abstract features serves as the model, and each sound is represented by the set of features that it possesses.

The spatial models are based primarily on measurements of proximity between stimuli, such as a direct judgment of how similar or dissimilar a pair of sounds is, the probability of one sound's being mistaken for another (confusion matrices), and so on. The fundamental assumption is that the measured proximity between two sounds has some systematic relationship to the geometric distance between the corresponding points. The primary technique for generating configurations of points from proximity data is a statistical technique called multidimensional scaling. Despite the fact that such representations are subject to some valid criticisms, they have been quite useful in practice. Their utility probably rests on two main points: (1) such representations can be suggestive and helpful even if imperfect and (2) multidimensional scaling is well developed and widely available. Further information about multidimensional scaling may be found in a variety of sources, such as Carroll and Kruskal (1978), Coxon and Davies (1982), Golledge and Rayner (1982), Green and Carmone (1970), Green and Rao (1972), Kruskal and Wish (1978), Law (1984), and Schiffman, Reynolds, and Young (1981).

The set-theoretic models are based on a wider variety of measurements. These include not only proximities like those used for spatial models, but also several other kinds of measurements, such as asymmetric judgments of similarity and dissimilarity (i.e., the responses to questions such as "How much is A like B?" and "How different is A from B?"). The fundamental assumption is that the measured value depends on three sets: the model features common to A and B, the features in A but not in B, and the features in B but not in A. In the best-developed version of this model (see Gati and Tversky, 1982), a count (possibly weighted) of the features in each of the three sets enters into a formula predicting the measured value. The formula used depends on the type of measurement. For example, if the data come from the question, "How much is A like B?," then the formula has the form:

$$w(\text{intersection count of } A + B) - u(\text{count of } A - B) - v(\text{count of } B - A)$$

where u, v, and w are positive and $u > v$. While models of this type are also subject to

some valid criticisms, they seem on the whole to permit a realistic representation of sounds. However, possibly because they are newer, and certainly because the methods for fitting them to data are little developed, their use is still limited.

Neither type of model should be taken seriously as describing how people or animals function. They both are capable of providing useful insights into such functions, but both surely fall far short of a description. It should also be noted that there is no conflict between them. There are certainly cases in which both can be used to good advantage on the same data, and they may even provide different sorts of information. Thus the use of each type of model should be based on its strengths and weaknesses and on what it has to offer in each situation.

CLASSIFICATION OF NONSPEECH TRANSIENT SOUNDS

Howard and his associates have undertaken studies focused directly on the classification of nonspeech transient sounds. Two sets of investigations are particularly relevant. In one series of studies, real-world and synthetic sounds were analyzed (using a crude auditory model) to determine which physical properties formed the basis of similarity judgments. In general, relatively crude properties, as might be conveyed by low-order principal components of spectral shape such as tilt and compactness, seem to correlate with the more important dimensions revealed by multidimensional scaling analysis of similarity ratings. Generally, listeners with musical training were more influenced by temporal properties of the sounds (periodicity and pitch) than by spectral shape properties, in contrast to listeners without such training. The second series of studies focused on the role of syntactic (structural) and semantic (interpretive) factors in determining the ability of listeners to distinguish specified sound patterns from randomly constructed patterns. Generally, listeners appear to utilize the syntactic structure provided by a simple finite-state grammar to improve the rate at which they learn to discriminate sound sequences and/or the final level of performance they can achieve. The effect of semantic themes is more problematic. For some listeners, instruction that provides thematic interpretation of sound patterns improves the discriminability of grammatical sequences; for others, it merely facilitates the initial learning of the discrimination task. Since reports on many of these studies have not yet been published, brief summaries of the studies are included below.

Howard (1976) asked 19 listeners (9 musically trained and 10 untrained) to rate the similarity of pairs of sounds drawn from a set of 8 underwater sounds. One-third octave spectra of these sounds were found to differ largely in terms of spectral compactness (phi-1) and spectral slope (phi-2). In addition, two of the sounds were distinguished by a low-frequency (under 1 Hz) periodicity. An INDSCAL analysis of the similarity ratings indicated that roughly 65 percent of the variance was attributable to three inferred dimensions that showed some correlation with the above three physical properties of the sounds. The similarity judgments of musically untrained listeners were more heavily influenced by spectral compactness, while those of musically trained listeners were more heavily influenced by periodicity.

Howard and Silverman (1976) asked listeners (11 musically trained and 23 untrained) to rate the similarity of pairs of sounds drawn from a set of 16 complex sounds. The physical properties of the sounds differed in a binary fashion across four dimensions: fundamental frequency (90 and 140 Hz), number of formants (one and two), formant frequency (low: 600, 1,550 Hz; high: 940, 2,440 Hz), and driving waveform (square and triangular). An INDSCAL analysis of the similarity ratings indicated that roughly 60 percent of the variance was attributable to three inferred dimensions. The first and second dimensions correlated with

fundamental frequency and waveform, while the third dimension correlated with formant frequency and number of formants. The similarity judgments of musically trained listeners (who were more homogeneous as a group with respect to feature saliency) emphasized fundamental frequency and deemphasized spectral shape, while those of musically untrained listeners emphasized spectral shape and deemphasized fundamental frequency.

In a large study (Silverman and Howard (1977), the authors measured listeners' ability to discriminate fundamental frequency, waveform, and formant frequency of 20 msec bursts of complex sounds followed (after a variable interstimulus interval—ISI) by a 500 msec burst of white noise. Performance was found to increase monotonically (with an exponential decay toward an asymptote) with ISI, with asymptotic value dependent on the size of physical difference to be discriminated and time constant (roughly 40 msec) that was independent of the property to be discriminated.

By constructing 16 noise signals with triangular envelopes differing in envelope periodicity (4–7 Hz) and attack/decay times (20 and 40 msec), Howard, Ballas, and Burgy (1978) had listeners rate the similarity of pairs of these and classify them into one of eight groups. An INDSCAL analysis of the similarity ratings indicated that roughly 69 percent of the variance was attributable to two inferred percent imensions that were correlated with envelope periodicity (tempo) and proportion of period spent in attack (quality). In all, 8 categories were used in the classification task, each category consisting of 2 of the 16 sounds. For the "tempo group," no two envelope rates were assigned to a single category; for the "quality group," no two attack times were assigned to a single category. Classification confusion matrices were analyzed using a model that assumed that the tempo and quality for each sound were uncorrelated Gaussian random variables with mean, but not variance, dependent on the corresponding physical parameter. As training (practice with feedback) progressed, the two variance parameters decreased for each group, but the decrease was more pronounced for the variance associated with the dimension along which it was necessary to make finer distinctions to achieve correct classification. For the tempo group, the variance approached the just noticeable difference (JND) for amplitude modulation rate in the frequency range used. In ancillary experiments, the assumption that tempo and quality were uncorrelated was verified, but similarity ratings of the category designations were similar across the two groups, thus indicating little effect of feature salience.

Howard and Ballas (1978b) asked listeners to rate the similarity of 16 tone complexes and also derived principal components of loudness compensated, one-third octave spectra of these sounds. The tone complexes consisted of 500 and 1,000 Hz components in four proportions superimposed on 22 inharmonic components (one-third octave spacing) with a Gaussian spectral envelope having one of four widths. The principal components analysis indicated that 91 percent of the spectral variance could be accounted for by two principal components, with the first component (74 percent) reflecting the average amplitude of components near the 500 and 1,000 Hz peaks, and the second component (17 percent) reflecting spectral slope. An ALSCAL analysis indicated that the similarity ratings could be accounted for by a two-dimensional solution (18.6 percent stress) with the dimensions corresponding roughly to the first two principal components of the physical spectra. The detailed clustering of points in the scaling solution, however, did not correspond well with the predictions of the principal components analysis: listeners appeared to dichotomize sounds along each dimension.

Several modifications of the work of Howard et al. (1978) were made by Ballas and Howard (1978a) in order to study classification. Two experienced and two naive listeners were tested. The range of variation of stimulus parameters was reduced: 4.8–6.4 Hz envelope rate (0.8 Hz steps) and 43–86 percent attack time (14 percent steps). Sound presentations

lasted 2.5 or 3.0 sec (rather than a fixed 3.0 sec) to discourage counting of cycles. More extensive practice with feedback was provided. Overall performance for the tempo partition was comparable to the previous study, but for the quality group performance after training was superior to that obtained previously. For the experienced listeners, the model variance associated with the stimulus parameter for which finer distinctions were required was smaller than that for the other parameter, as in the earlier study. Relative to the previous study there was less difference between the accuracy for tempo and quality when the classification stressed quality than when the classification stressed tempo, a result the authors attribute to reduced discriminability for quality differences.

Howard and Ballas (1980) studied how well subjects could learn to discriminate a set of sound sequences generated by a finite-state grammar from randomly generated sequences relative to arbitrarily selected random sequences. The sounds consisted of 80-msec tone bursts (1,157, 1,250, 1,345, 1,442, and 1,542 Hz), 82 msec of unrelated real-world transients, and 320-msec sounds related to water and steam. The finite grammar had six states and, in the case of the tone burst sequences, constrained the initial two sounds to either 1,157- or 1,250-Hz bursts and the final sound to either a 1,442- or 1,542-Hz burst. Learning appeared to be faster for the grammatically generated sequences than for the random sequences for all sets of sounds. Furthermore, there appeared to be substantial generalization to unfamiliar grammatical sequences, but not, of course, to unfamiliar randomly generated sequences. The ability to recoginize grammatical sequences of related sounds was somewhat improved when subjects were given instructions suggesting semantic interpretations for the sounds. The investigators interpreted these results as indicating that both syntactic and semantic factors can play important roles in the classification of acoustic transient patterns.

Howard and Ballas (1981) studied various ways of training subjects to detect grammatical patterns of real-world sounds (or visually presented verbal descriptions of the sounds) related to water and steam. In the main experiment, training consisted of either practice in classification (with feedback) or observation of the patterns (without feedback). The results indicate that observation alone improves initial classification performance, that visual observation of verbal descriptions is as effective as listening to sound sequences in training classification, and that asymptotic performance is the same for all groups (independent of classification task or training technique). In ancillary experiments, training classification of real-world sounds by observation tone sequences with the same grammatical properties was found to be relatively ineffective, but the monotonicity of the mapping from observation tones to classification tones was found to have little effect on classification performance.

Howard and Ballas (1981) studied simultaneous detection and identification of grammatical sequences of tone pulses (150 msec, 100 msec IPI, five frequencies ranging from 1,000 to 1,500 Hz with 125 Hz spacing) using different types of training. Twelve grammatical and nongrammatical sequences were designated targets for both detection and identification. In the first experiment, training consisted of practice on the task with feedback. Detection performance reach an asymptote rapidly to a near-perfect level for both types of targets, but identification performance improved slowly, with asymptotic performance on the nongrammatical targets better than for the grammatical targets. In the second experiment, training consisted of target observation with feedback of target identity. Performance did not improve during the testing (with feedback) conducted post-training, and performance for the grammatical and nongrammatical sound patterns was essentially identical, although somewhat below that observed in the first experiment. In the third experiment, observation and testing (without feedback) were interleaved. Performance on both the detection and identification tasks was more accurate for the grammatical sequences. A comparison with

previous studies of detection alone indicated that existence of a simultaneous identification task improved detection performance.

The effect of syntactical structure in the detectability of sequences of real-world sounds (related to water and steam) was investigated by Howard and Ballas (1980). In the main experiment, targets were either grammatical or nongrammatical sequences, and half the subjects read a 30-word description of the sounds. The results indicate that detection performance for the grammatical sequences was superior to that for the nongrammatical sequences. However, the effect of the verbal description was restricted to improving initial detection performance on the grammatical sequences. In a second experiment, the identities of the components of the grammatical target patterns were permuted to make the sequences more difficult to interpret. For these sequences, asymptotic detection performance was superior for listeners who had not received verbal descriptions of the sounds. The results were interpreted as indicating that, although both sequential structure and semantic factors can play a role in nonspeech pattern classification, structure is more important than interpretability in determining detection performance.

SONAR DETECTION BY HUMAN OBSERVERS

The sonar operator's task is to detect and classify signals received from the underwater sound environment. A number of studies have attempted to define how the processing of these complex nonspeech sounds depends on the physical properties of the signals and on the listener's training, knowledge, and expectations. In general, the phenomena revealed by these studies are quite similar to those observed in experiments on auditory psychophysics and speech perception.

Some experiments have addressed the possible deleterious effects of prolonged watch periods on sonar monitoring. Contrary to earlier data, O'Hanlon, Schmidt, and Baker (1965) found no impairment in a listener's ability to detect doppler shifts (small frequency changes) after prolonged listening to sonar returns. Kobus et al. (1986) studied the detection and recognition of simulated sonar targets using simultaneous auditory and visual modes as well as both modes alone. They reported no advantage for dual mode over single mode performance, contradicting a previous study by Colquhoun (1975).

Several investigators have been concerned with how sonar operators identify waterborne noises. Corcoran et al. (1968) reported several factors that could improve training for sound identification: the use of verbal labels, feedback, specific stimulus orders, and signal-to-noise ratios. Webster, Woodhead, and Carpenter (1973) studied the discrimination of 16 speechlike and enginelike sounds. These sounds took binary values on four dimensions: (1) source harmonic structure, (2) fundamental frequency, (3) number of formants, and (4) formant frequencies. Fewer confusions were made between sounds with a greater number of differing dimensions; the relative importance of the dimensions decreased from (3) to (4) to (2) to (1). Subjects seemed to weight heavily the complexity and periodicity of the signal.

A number of studies have applied multidimensional scaling techniques to the perception of sonar signals. In a scaling analysis using the Webster et al. (1973) signals, Morgan, Woodhead, and Webster (1976) successfully recovered the signals' known structure. Other scaling studies have been performed by Howard and Silverman (1976) using a similar 16-signal set, by Howard (1977) using an 8-signal set, and by Mackie et al. (1981) using Howard's (1977) set as well as larger sets of actual underwater signals and experienced sonar operators as listeners. The resulting similarity spaces depend on the set of signals used and on the listener's training and experience. Howard and Ballas (1981, 1983) and Howard (1982) proposed that the listener's perceptual space reflects contextual properties of the signal

set. In Howard's (1982) model, a low-resolution spectral analysis (third-octave filtering) is followed by a principal-component analysis of the signal ensemble. Their experimental results supported the assumption that listeners' use of signal features is dependent on the task context.

Howard and Ballas (1980, 1982) demonstrated that higher-level factors can influence the classification of nonspeech transient patterns. They trained observers to classify sequentially structured patterns of complex sounds (clank, flush, etc.) as either targets or nontargets. Syntactic factors (a target generated by a defined finite-state grammar) and semantic factors (lifelike source event sequences) produced the expected effects on target classification.

4
Auditory Object Perception

The text by Moore (1982) provides a useful introduction to the topic of auditory objects and patterns. Moore organizes the subject in terms of: (1) object perception and identification, (2) separating objects, (3) perception of temporal patterns, and (4) general principles of perceptual organization.

OBJECT PERCEPTION AND IDENTIFICATION

For sounds consisting of a single frequency, two numbers are sufficient for classification: frequency (pitch) and/or intensity (loudness). Humans can identify only 5–6 simple sounds out of large set of tones varying in either frequency or intensity (Pollack, 1952). As a sound's spectral complexity (number of frequency components in the sound) increases, the number of possible classifications also increases. For complex sounds, the dimension of spectral complexity, sometimes associated with the percept of timbre, is used to describe the sound. The spectral complexity, and thus the timbre, can be static or dynamic over time. In time-varying patterns, the onsets and offsets of the sound, especially in music, play a crucial role in sound identification.

The steady-state spectrum is important to listeners as they describe sounds along one of the timberal dimensions, for example as mellow or brilliant (von Bismarck, 1974a, 1974b). Scaling studies based on judgments of similarity of complex tones show that spectral distribution of energy is a major factor (Wedin and Goude, 1972; Plomp, 1976; Grey, 1977). The significance of this dimension was confirmed by Grey and Gordon (1978), who exchanged spectral envelopes among their stimuli and observed an exchange of positions along the axis assigned to steady-state spectra. A second major factor, revealed by judgments of similarity among the tones of musical instruments, is the synchrony among harmonics during attacks or other temporal fluctuations (Wessel, 1979), and a third appears to relate to the presence of high-frequency energy, probably noise, during attack (Grey, 1977).

Less work has been done on the matter of identification. There is a widely held opinion that the fine details of the steady-state spectrum cannot play a major role in identification. Sound sources, for instance different musical instruments, can be successfully identified under diverse listening conditions that markedly distort the steady-state spectrum. Outside the speech domain, however, there is little quantitative work on the nature of spectral distortion that would actually impair identification. Berger (1964) showed that if musical instrument tones are low-pass-filtered so that only the fundamentals survive, then

identification performance is dramatically reduced. In normal conditions, time-varying effects, particularly onset transients, may be more important than steady-state spectrum in identification. There is circumstantial evidence to favor this view in the work of Grey (1977) and his colleagues: the tones from musical instruments of the same family tended to cluster along dimensions associated with transient temporal features of the signals. It should be noted, however, that the stimuli used in these studies were of such brief duration that identification was nearly impossible.

SEPARATING OBJECTS

Separating objects refers to the ability of listeners to separate perceptually, and to identify, simultaneously sounding sources, especially when the spectra of these sources are interleaved. It is evident that this ability cannot be tonotopically based. Reviews by McAdams (1984a, 1984b) and by Hartmann (1987) make note of a number of signal characteristics that affect object separation: spectral profile, temporal modulation, onset/offset characteristics, and spatial separation.

Spectral Profile

In the context of object perception, spectral profile refers to the arrangement of the spectral components that make up a complex sound. Clearly, if a set of components is much greater in amplitude than the other components of a sound, then the more intense components are likely to form an auditory object (McAdams, 1984a). Changes in the spectral location of the components may also significantly alter the perception of the sound. An obvious example is that altering the spacing of harmonics of sound will lead to a change in the sound's pitch and/or timbre.

Increasing the number of spectral components makes it more difficult to hear individual components (Plomp, 1964, 1976; Plomp and Mimpin, 1968) and promotes synthetic listening, as in the work of Patterson (1973). A peak in the spectral envelope promotes a separation of a component at the peak (Martens, 1981). Harmonics of a complex tone with high harmonic numbers can be separated more readily than those with low harmonic numbers (Houtsma, 1981). This result appears paradoxical from a tonotopic point of view. One would expect that a spectral envelope that decreases with increasing frequency should promote fusion among the partials, although experiments by Martens (1981) and McAdams (1984a) do not support this conjecture.

The work of Green and his colleagues (see Green, 1988, for a review of this work) has demonstrated the importance of the contour of amplitudes in the spectrum of a complex sound for discriminating among stimuli with different spectral profiles. Sounds with subtle changes in the amplitude profile can be discriminated despite large random variations in the overall amplitude of the sound.

The various models of complex, or virtual (Terhardt et al., 1982), pitch are based on various forms of spectral pattern recognition (see de Boer, 1976, for a review). The spacing of the components in a complex spectrum is a major determinant of the pitch and, to some extent, the timbre of the sound. This empirical and theoretical work indicates that small differences in the amplitudes of spectral components and in the spacing of the components of a complex sound may be a basis for classification.

Temporal Modulation

Modulation Defined

There is a class of complex signals that may be called "modulated." As commonly used, the term ***modulation*** refers to a periodic or nearly periodic variation with time of some parameter of an acoustical signal, for example the amplitude (AM) or the frequency (FM). Restricting the definition of modulation to periodic or nearly periodic variations has the advantage of extending the concept of the steady state; a modulated signal maintains all of its physical character, a deterministic character, indefinitely. It has the disadvantage of excluding nonrepetitive variations that might be comprehended with perceptual models similar to those used for modulation perception.

Variations that are not classified as modulation because they occur only once during a time interval of interest may be called "transient." Variations that are not classified as modulation because they are random may be called "fluctuations," although in the case of noisy variations the stochastic character is normally maintained indefinitely.

Modulation is present in nature, for example, in bird calls (Greenewalt, 1968) and in music as tremolo or vibrato (Seashore, 1932, 1935). Modulation is present in the sounds of virtually any machine in which there is a rotating element.

By far the majority of the work done on the perceptual effects of modulation has been concerned with the detection of modulation. The impetus for modern work (post World War II) was the 1952 study by Zwicker on FM and AM detection, as a function of modulation frequency. Zwicker's study shifted the emphasis from an exclusive concern with the connection to difference limens (e.g., Riesz, 1928; Shower and Biddulph, 1931) to a tonotopic reference and measures of the critical band. Zwicker opened the question, which has yet to be fully resolved, as to whether AM and FM detection can be understood from a common perceptual model (e.g., Maiwald, 1967a, 1967b, 1967c, and Goldstein, 1967) or whether separate perceptual processes must be involved. Recent research, for example that of Coninx (1978) and Demany (1985), supports the latter view.

As elsewhere in psychoacoustics, the matter of modulation detection can be approached from either spectral or temporal points of view. But for modulation detection, there is at least some guideline based on the modulation frequency. The case of high modulation frequencies (greater than half the critical bandwidth at the carrier frequency) can be considered a solved problem. The correct approach is spectral, and modulation detection is equivalent to a masked threshold (Hartmann and Hnath, 1982; Schorer, 1986). At low modulation frequencies, at which modulation detection might be regarded as an alternative to discrimination, the temporal point of view seems most attractive (Hartmann and Klein, 1980), although there is evidence from Fastl (1978) and from Demany and Semel (personal communication, 1987) that the Hartmann-Klein model may fail at high carrier frequencies.

An alternative to the temporal approach at low modulation frequencies is the suggestion by Kay and Matthews (1972) that modulation (specifically FM) is detected in channels tuned to specific modulation frequencies. This suggestion, based on data from selective adaptation experiments, has never been adequately confirmed or rejected.

Suprathreshold Modulation Perception

Studies of suprathreshold modulation perception appear to have been limited to investigations of frequency modulation by different complex waveforms. Divenyi and Hirsh (1972) compared the sensations elicited by triangle, trapezoidal, and square-wave modulation. Of interest were the relative frequency excursions required to produce the same sensation of

modulation width. Such points of subjective equality (PSE) occurred, for example, when triangle modulation had an amplitude 1.72 times greater than square-wave modulation. PSEs for other waveform combinations were found by Hartmann and Long (1976) and by Hartmann (1985). Klein (1980) found PSEs for sine modulation compared with a modulation waveform comprised of first and third harmonics. The results showed that PSEs depend on the relative phases of the first and third harmonics, which is evidence against the Kay-Matthews channels hypothesis. The results also showed that PSEs depend on the physical width of the standard, an observation that excludes models based entirely on scaling and somewhat complicates attempts to understand the perceptual process.

An amplitude modulation imparted identically to two sources can cause fusion of dichotically presented sounds (von Bekesy, 1963) or of inharmonic sounds (Bregman, Abramson, and Darwin, 1985). The role of frequency modulation has been explored by Chowning (1980) and by McAdams (1984a). Fusion among spectral components or groups of spectral components is promoted by a common FM; separability is promoted by giving components, or groups of components, different FM waveforms. The study by McAdams extended the FM technique to include jitter, small random frequency fluctuations that are present in the sounds of all musical instruments, whether played with vibrato or not.

Spectral Interaction Among Stimuli That Are Temporally Modulated

Recent work with a variety of stimuli that have slow temporal modulation patterns, especially amplitude-modulated patterns, has demonstrated an interaction among frequencies that lie outside the traditional estimates of the critical band of the signal being processed. The work on comodulation masking release (see Hall, 1987, for a review) shows that the detection of a tonal signal masked by a narrow band of noise can be improved by as much as 10–12 dB if a band of noise with the same temporal modulation as that of the masker is presented in a spectral region outside the critical band containing the signal. The correlation or comodulation between the two noises appears to be the important factor in aiding the detection of the masked signal. Other research (Cohen and Schubert, 1987; McFadden, 1987; Wakefield and Viemeister, 1975; Yost and Sheft, 1988) has shown that under some conditions the interaction of two amplitude-modulated signals in different spectral regions may interfere with a listener's ability to process the target signal. Yost and Sheft (1988) suggest that these interactions may be a consequence of the auditory system operating as a wide band detector in order to find common patterns of temporal modulation across the spectrum of a complex sound. As discussed above, these temporal patterns may aid the system in identifying auditory objects.

Onset/Offset Characteristics

Onset asynchrony dramatically increases separability even though the asynchrony may not be otherwise apparent (Rasch, 1978, 1979). The work of Summerfield et al. (1987) and the Kubovy and Jordan (1979) phase-shift experiment similarly emphasize the significance of temporal changes. Inharmonicity among the spectral components promotes separability (Martens, 1984). Models of separation based on inharmonicity have been constructed by Duifhuis, Willems, and Sluyter (1982), Terhardt et al. (1982), and Scheffers (1983). Similarly, a common attack and decay envelope aids in the fusion of inharmonic partials (Mathews and Pierce, 1980; Cohen, 1984).

Decreasing the duration of a signal promotes fusion (Moore, Peters, and Glasberg,

1985a; Hartmann, 1985). Even inharmonic tones are fused if they are brief enough. Conversely, the partials of a steady harmonic tone can be heard if its duration is long enough (Helmholtz, 1855). Separation of objects requires information, whereas fusion appears to be the default percept. As might be expected then, musical training promotes analytic listening, in which partials are separated (Soderquist, 1970; Houtsma, 1979). Fusion, or synthetic listening, is promoted by low sound pressure levels, at least in the context of experiments in which the listener's task requires the synthesis of a low pitch (Houtsma, 1979).

Spatial Separation

One way to classify a sound is to place it at some location in auditory space. The sound's attribute is then a spatial coordinate. Spatial location is also one way to separate one sound source from other sound sources. The acoustic variables that determine a sound's source have been investigated for hundreds of years. Auditory sensitivity to the two basic binaural cues, interaural time and level (the duplex theory of localization, see Stevens and Newman, 1936), has been discussed in many excellent review articles and books over the past two decades (Green and Henning, 1969; Mills, 1972; Durlach and Colburn, 1978; Blauert, 1982; Gatehouse, 1985; Libby, 1980; Yost and Gourevitch, 1987). More recently spectral cues, both monaural and binaural (Butler, 1985; Blauert, 1982; Hartmann, 1983), have been identified as major variables for complex sound localization. The changes that the spectrum of a complex sound undergoes from its source to the inner ear, especially at the head, torso, and pinna (Kuhn, 1987; Blauert, 1982; Butler, 1975; Wightman, Kistler, and Perkins, 1987) are crucial transformations for determining the source of sound, especially if the sound has a high-frequency spectrum.

Localization in complex acoustic environments has also received considerable attention, especially with regard to localization in rooms (see review chapter by Berkely, 1987). Architectural acousticians have studied the effects of room reverberation and absorption on the ability of listeners to locate sounds in enclosed spaces. Alterations of both the spectrum and the time domain of a waveform take place in an enclosed space. These changes can alter the quality of the sound source (i.e., coloration, see Yost, 1982; Yost, Harder, and Dye, 1987) and its apparent location (i.e., precedence, see Zurek, 1987).

Spatial separation promotes the perceptual separation of auditory objects, as shown in the cocktail-party effect experiments by Cherry (1953). Dichotic pitch phenomena may be regarded as the separation of a tone from a noise background, on the basis of an interaural time difference as a spatial cue (Cramer and Huggins, 1958; Bilsen and Goldstein, 1974; Klein and Hartmann, 1981). But spatial separation by no means guarantees perceptual separation of the objects. The octave illusions of Deutsch (1974) depend on fusion of a dichotically presented tone, as does the dichotic periodicity pitch of Houtsma and Goldstein (1972).

The cocktail-party effect (see Cherry, 1953; Cherry and Wiley, 1967) refers to the presumed ability to use binaural cues to easily recognize a particular sound source in a noisy environment. That is, a complex signal can be extracted from a noisy environment better when two ears are used than when one ear is used (Cherry, 1953). This enhanced recognition ability is presumably due to the binaural system separating the signal of interest from the noise background when the spatial location of the signal is different from the rest of the background sounds.

Studies of the binaural masking-level difference (BMLD or MLD, see Green and Yost, 1975; McFadden, 1975; Colburn and Durlach, 1978; Durlach and Colburn, 1978; Jeffress,

1972; and Durlach, 1972, for reviews) demonstrate the advantage for detection of a signal, presented over headphones, with a different interaural configuration than that of a masker. Similar detection advantages exist when a signal to be detected is presented from one loudspeaker and a masking stimulus presented from another loudspeaker (Plomp and Mimpin, 1981). Most models of the MLD are based on the ability of the binaural system to process interaural differences of time and level, and as such these models are functionally equivalent to localization models (Colburn and Durlach, 1978). Binaural advantages for discrimination or identification of sounds are much smaller or nonexistent compared with the detection results (see Green and Yost, 1975) described above.

The literature on auditory streaming clearly shows that auditory space can be used to separate one sound group from other groups (Bregman, 1978a; McAdams, 1984a). The phenomenon is more striking over headphones than over loudspeakers, perhaps because greater interaural differences can be presented over headphones. Kubovy (1987) argues that although space can be used to separate sounds, frequency or pitch is a more potent cue for segregation. Kubovy states that this is because the auditory system, unlike the visual system, is tonotopically structured, not spatiotopically organized. The basic transformation in hearing is from frequency to neural location, while for vision (and on the skin) the basic transformation is from space to neural location.

Consideration of auditory localization reveals a remarkable ability of the auditory system. When a complex sound moves through space (e.g., a person walking through a room), the sound source undergoes numerous physical and physiological transformations. Yet listeners, in all but the most unusual conditions, perceive a fully integrated acoustic image moving continuously through space. The way in which the nervous system separates the sound source from the other sounds and determines its location should provide valuable insights concerning the auditory system's ability to classify sounds in general.

PERCEPTION OF TEMPORAL PATTERNS

Streaming

The concept of streaming concerns the tendency for certain sequences of sounds in a complex sound field to appear as one object, as if this sequence were a stream isolated from other sounds (Bregman, 1978a). Sound sequences with spectral, spatial, intensive, and temporal similarity often form such streams. However, the most powerful cue for stream segregation is similarity in the spectral dimension.

Patterns of tones may be perceived as a single stream or as segregated streams (van Noorden, 1975; Bregman and Campbell, 1971). In the case of segregated streams, the percept of temporal order across streams is virtually lost. Stream segregation experiments typically use the frequency range as the major parameter. Dowling (1968, 1973) has reported that streams may be segregated on the basis of intensity or spatial location, but van Noorden (1975) finds intensity-based streaming to be relatively weak.

It is reasonable to conjecture that streaming is related to object separation (Bregman 1978a): auditory objects are interpreted as sources and successive sounds from a single source form a stream. Experimentally, however, the picture is not so clear. Most demonstrations of stream segregation have been tonotopically based, with pitch range as the major streaming parameter. Wessel's (1979) demonstration of streaming by timbre is also tonotopically based, with spectral envelope playing the major role. To demonstrate a close association between stream segregation and object recognition would require evidence that stimulus factors that affect object separation and recognition (steady-state spectrum,

transient character, or factors from the above list) operate similarly on stream segregation. Some data along these lines have been collected by Bregman (1978b).

Perceptual Restoration of Masked Sounds

Since we live in a noisy world, signals of importance are often accompanied by extraneous sounds that mask fragments of these signals. In recent years, it has been recognized that we possess a rather sophisticated series of mechanisms for reversing the effects of masking through perceptual synthesis of obliterated portions of sounds of interest. As we explain below, this restoration is based on contextual information furnished by preceding and following segments of the obliterated portion, as well as an analysis that ensures that the interfering sound has spectral components of an appropriate amplitude capable of masking the sound that is restored.

The restoration of obliterated sounds is known as "auditory induction." In the laboratory, obliteration can be accomplished in two ways–either by adding a masker to the signal or by deleting the signal and filling the gap with a louder sound. The latter method is preferred by most investigators since it ensures complete masking.

Three types of auditory induction deal with the restoration of obliterated fragments of signals: (1) Heterophonic continuity involves the illusory continuation of one sound when interrupted by a different louder sound; (2) homophonic continuity is the illusory continuity of a sound when interrupted by a louder level of the same sound; (3) contextual concatenation, which does not involve illusory continuity of a steady-state signal as do the other types of auditory induction, consists of restoration of an item that differs from the preceding and following sounds. An especially interesting type of contextual concatenation is phonemic restoration, in which speech segments are restored in keeping with the application of syntactic and semantic rules. These three types of auditory induction follow the same acoustically based rules, as we discuss below.

Heterophonic Continuity

The illusory continuity of one sound when interrupted by a louder sound has been discovered independently several times. The first discovery was that of Miller and Licklider (1950), who found that a tone was reported as being on all the time when it was alternated with a louder broad-band noise, each sound lasting 50 msec. They compared illusory continuity to gazing at a landscape through a picket fence: In spite of the interruptions, a viewer considers the background to be continuous behind the pickets. Vicario (1960) rediscovered the illusory continuity of a sound interrupted by a noise, which he called the acoustic tunnel effect. He considered the illusion to be analogous to the visual tunnel effect, a phenomenon studied by Gestalt psychologists who noted the apparent presence of an object when it moved behind a closer opaque body. Thurlow (1957) was responsible for another independent discovery of heterophonic continuity. He alternated two tones (each lasting 60 msec) that differed both in frequency and intensity and observed that the fainter tone appeared to be continuous. He considered the illusion to be an auditory analog of the visual figure-ground effect, in which contours are perceived as part of a visual figure, while the background is considered to be present behind the figure. This work by Thurlow was the basis for a number of experiments in which heterophonic continuity was studied for durations ranging from 10 through 100 msec (Elfner, 1969, 1971; Elfner and Caskey, 1965; Elfner and Homick, 1966, 1967a, 1967b; Thurlow and Elfner, 1959; Thurlow and Marten, 1962). Using the results of these studies, Thurlow and Erchul (1978) developed the theory

that illusory continuity was the consequence of a continuation of a firing of neural units corresponding to the fainter sound as a result of facilitation produced by the louder sound. This was an extension of a similar hypothesis made earlier by Thurlow and Elfner (1959). Thurlow and Erchul mention the possibility that this facilitation of prior activity might result from an excitatory postsynaptic potential. This model for heterophonic continuity does not require that the louder of the sounds be capable of stimulating directly the auditory units stimulated by the fainter sound, as is required by the subsequent models discussed below.

Houtgast (1972) considered that illusory continuity of tones interrupted by louder sounds could be used to study peripheral events leading to perception. He alternated tones with durations of 125 msec with louder sounds of equal duration and appropriate spectrum and intensity. He measured the level at which the discontinuity of the tone could be detected, calling this value the pulsation threshold. Houtgast suggested that the following rule determined the level of this threshold: "When a tone and a stimulus S are alternated (alternation cycle about 4 Hz), the tone is perceived as being continuous when the transition from S to tone causes no (perceptible) increase of nervous activity in any frequency region." This rule provided a neural basis for quantitative psychophysical measurements and resulted in the use of pulsation thresholds to study peripheral events leading to stimulation of the auditory nerve. Among the topics studied using this technique are the shape of psychophysical tuning curves (and their relation to neurophysiological tuning curves), the width of critical bands (a measure of the frequency resolution of the cochlea), and the extent of lateral suppression (the reduction of neural sensitivity at the edges of stimulated regions) (see Aldrich and Barry, 1980; Fastl, 1975; Glasberg, Moore, and Nimmo-Smith, 1984; Houtgast, 1972, 1973, 1974a, 1974b; Kronberg, Mellert, and Schreiner, 1974; Shannon and Houtgast, 1986; Verschuure, Rodenburg, and Maas, 1974; Weber, 1983). This procedure is not without its critics. Bregman and Dannenbring (1977) questioned the concept that continuation of neural activity was required for perceptual continuity. They alternated a tonal signal with a noise-producing auditory induction and introduced an intensity ramp that increased the intensity of the tone just before the onset of the louder noise, reasoning that "turning up the tone just before the noise might boost the neural activity corresponding to the tone and increase the illusion of continuity" (p. 157). They found that illusory continuity was prevented by presence of the ramp and concluded that this finding was not consistent with Thurlow and Elfner's (1959) neurofacilitation model, Houtgast's (1972) neurocontinuity model, and other variants of these models. However, Bregman and Dannenbring's observations are consistent with the contextually driven restoration mechanism described below.

Warren, Obusek, and Ackroff (1972:1151) proposed the following rule for temporal induction: "If there is contextual evidence that a sound may be present at a given time, and if the peripheral units stimulated by a louder sound include those which would be stimulated by the anticipated fainter sound, then the fainter sound may be heard as present." This rule was somewhat broader than Houtgast's for continuation of steady-state signals as cited above. This extended coverage was designed to encompass phonemic restorations of obliterated segments of speech, which had been discovered a few years earlier (Warren, 1970). In the main experiment of Warren et al. (1972), they alternated an 80 dB, 300 ms, 1,000 Hz pure tone (the inducer) with fainter 300 ms pure tones ranging in frequency from 150 Hz through 8,000 Hz. The intensity limits for illusory continuity of the fainter tones were determined and compared with simultaneous masking functions. When the 1,000 Hz tone at 80 dB remained on continuously, and the masked threshold was determined for superimposed intermittent tones on for 300 ms and off for 300 ms using the same tonal

frequencies employed for auditory induction measurements, the correspondence between masking functions and induction functions met the requirements of theory and has been verified by subsequent studies.

The maximum duration of illusory continuity of tones induced by other tones or by noise is about 300 ms (Verschuure, 1978). However, remarkably long durations of illusory continuity were reported for narrow band noise induced by a louder broader band noise, with a fainter noise seeming to continue along with the louder noise for some tens of seconds (Warren et al., 1972). There is no explanation currently available for this extraordinarily long-duration continuity of narrow band noise. Most studies of temporal induction have used the signal intensity limits as the dependent variable, and it would be of interest to systematically study factors influencing durational limits.

Homophonic Continuity

Homophonic continuity is the simplest type of auditory induction. Its special interest lies in the insight it provides concerning the manner in which the inducer enters into the perceptual synthesis of the fainter sound. Homophonic continuity is produced when two levels of the same sound are alternated (Warren et al., 1972). Two levels of any sound can be used, but let us consider the case of a 300 ms broad band noise at 80 dB alternated with 300 ms of a 65 dB level of the same sound. The 65 dB level will appear to be on continuously with the pulsed addition of a louder level. This illusory continuity is in a way paradoxical, since the fainter sound would be masked completely were it present along with the louder sound. Homophonic continuity can be used to illustrate the subtractive nature of auditory induction. When noise at 70 dB is alternated with the same noise at 72 dB, then the 70 dB level appears continuous with the pulsed addition of a fainter sound. The fact that the 72 dB inducer seems fainter than the 70 dB continuous sound can be attributed to the fact that if the 70 dB level is subtracted from 72 dB, the residue is less than 70 dB (in fact, 67.7 dB), and it is this residue that is heard as a pulsed addition to the continuous level. While the subtractive process is especially easy to demonstrate with homophonic induction, an analogous procedure of subtracting neural activity corresponding to the restored sound from the inducer appears to occur both for heterophonic continuity and for contextual catenation.

Contextual Catenation

Homophonic and heterophonic continuity both involve restoration of segments of a continuing steady-state sound. When the stimulus that is interrupted is one that changes with time, then the obliterated fragment differs from the sounds that immediately precede and follow the interruption. A more complex type of perceptual synthesis is required under this situation. This restoration (which is called contextual catenation) involves using situational information. A simple type of contextual catenation was described by Dannenbring (1976), who interrupted tonal glides with a louder broad band noise. He reported that, under appropriate conditions, the tone was heard to continue its glide through the broad band noise for intervals of a few hundred milliseconds. In keeping with the rules governing other types of auditory induction, the inducing noise needed to be capable of masking the restored tonal glide had it been actually present along with the noise. A somewhat different type of tonal extrapolation was reported by Sasaki (1980), who found that the illusory perception of missing notes of a familiar melody played on a piano were restored when these notes were replaced by a louder noise. The most complex (and the

most thoroughly investigated) type of contextual catenation is that occurring with speech. These phonemic restorations involve the application of syntactic and semantic rules to help identify the obliterated fragments.

In the first study dealing with phonemic restorations, the segment of speech indicated by the asterisk (together with portions of the preceding and following speech sounds) was removed completely from a recording of the sentence, "The state governors met with their respective legi(*)latures convening in the capital city," and the deleted portion was replaced by a louder cough having the same duration. Listeners believed that the sentence was intact with no speech sound missing, and they could not locate the position of the cough. When told that a portion of the sentence had been deleted and a cough substituted, the listeners still could not tell which portion of the sentence was missing even after listening to the recording several times (Warren, 1970; Warren and Obusek, 1971). The cough appeared to occur along with the sentence, but appeared to float alongside without any recognizable position. Phonemic restorations were also induced by other loud sounds, but when a blank piece of tape having a duration equal to the deleted segment was spliced into the sentence, the location of the silent gap could be identified and listeners could tell which of the speech sounds was missing. Restoration was not limited to single phonemes, and entire syllables could be restored. Sasaki (1980) studied phonemic restorations in Japanese speech and reported that individual phonemes and entire syllables could be restored perceptually when deleted and replaced by noise.

The contextual catenation of phonemic restorations can involve the use of subsequent as well as prior information. It has been reported that when the identity of the deleted speech sound is ambiguous on the basis of earlier portions of the sentence, but resolved by information following the deleted segment, this later information can be utilized to determine the nature of the restoration (Warren and Warren, 1970).

Samuel has reported several studies dealing with phonemic restorations. He found that restoration of a particular speech sound was enhanced when the replacing sound was acoustically similar to the deleted phoneme (Samuel, 1981a). In another study, Samuel (1981b) superimposed the extraneous sound on the speech sounds and required listeners to report whether the utterance was intact (with an added extraneous sound) or had a portion removed. Using a signal detection methodology, he obtained a miss rate and a false alarm rate and then calculated the parameters of discriminability and bias under a variety of conditions. Samuel and Ressler (1986) considered that configurational properties of a word could interfere with attention to individual phonemes and thus enhance phonemic restorations. They trained subjects to process individual phonemes in a word selectively. With some trials they also provided a visual prime that served as an attentional cue. They found that the priming that identified both the word and the phoneme could inhibit phonemic restorations.

Perception of Acoustic Sequential Patterns

Perception of temporal order has been a topic of considerable interest, due largely to the fact that speech and music consist of ordered sequences of sounds. An ecological approach to the perception of temporal order involves the determination of the rate at which successive sounds occur in speech and music. Fraisse (1963) stated that the fastest rate of successive notes found in concert selections corresponds to about 7 per second, or 150 msec per note. However, it seems that familiar melodies can still be recognized down to about 50 msec per note (Winkel, 1967). Winkel noted that some composers used more rapid rates, but that such ornamental playing was heard as "flickering" or

"rustling." The rate of phonemes in speech is somewhat more rapid than the notes in music. Conversational English has about 120 to 150 words per minute, and, considering the average word to contain five phonemes, the average duration per phoneme has been calculated to be about 80 to 100 msec (Efron, 1963). Speech when read aloud is more rapid than conversation, and for reading the average duration drops to about 80 msec per phoneme. It was stated by Joos (1948) that intelligibility is reduced when the average duration of phonemes reaches 50 msec. With some practice, "compressed speech" (recorded speech that is accelerated by special devices while keeping pitch constant), retains some intelligibility when the average duration of phonemes is only about 30 msec (Foulke and Sticht, 1969). Early work by Hirsh (1959) and Hirsh and Sherrick (1961) established that with pairs of sounds consisting of tones, hisses, and clicks, order could be identified with onset differences of the sounds down to about 20 msec. Broadbent and Ladefoged (1959) claimed that at such brief durations, discrimination of order was accomplished indirectly through recognition of qualitative differences in the sound pairs, a conclusion that was contested by Hirsh and Sherrick. Further systematic work with two-item sequences was reported by Kinney (1961) and by Fay (1966). Two-item sequences were used with aphasics by Efron (1963) and Tallal and Piercy (1974) to determine if problems in distinguishing orders were associated with speech disorders (they were). Other experiments were reported for two-item sequences in which listeners were required to discriminate between different orders without naming. Under these conditions very low thresholds were reported that were associated with pitch-quality differences. Patterson and Green (1970) used pairs of brief clicklike sounds (called Huffman sequences) having identical power spectra but different phase spectra, so that the only difference between members of a pair was in temporal arrangement. They found that Huffman sequences permitted discrimination between temporal orders down to 2.5 msec. Yund and Efron (1974) found that listeners could discriminate between permuted orders of a two-item sequence (such as two tones of different frequencies) down to temporal separations of only 1 or 2 msec. Wier and Green (1975) reported similar results for patterns of two tones with a total duration of only 2 msec. Efron (1973) emphasized that such micro patterns were perceived as unitary perceptual events, with different qualities associated with the different orders. Listeners could not identify the order of components within these brief sequences unless such information was given to them. Efron pointed out that once subjects had learned the temporal order corresponding to the characteristic quality of the stimulus pair, they could infer the correct order on subsequent presentation.

In the early 1970s, experiments were reported that indicated that the initial and terminal items of sequences are identified with special ease, so that results obtained with two-item sequences could not be generalized to sequences containing more items (Warren, 1972; Divenyi and Hirsh, 1974). Pastore (1983) reported that thresholds for identification of temporal order were shorter for offset asynchronies than for onset asynchronies, an effect he attributed to the greater availability of echoic information in the offset condition. However, three procedures have been developed that deal with this problem of special onset and offset cues: (1) use of complex multielement sequences consisting of 10 or more items, (2) use of extended binary sequences, and (3) use of repeated or looped sequences of a few sounds (three to six sounds repeated over and over without pauses).

Complex Multielement Sequences

Watson and his coworkers generally employed "word-length" sequences of 10 tones having frequencies within the range important for speech (300 to 3,000 Hz) and durations of 40 msec, approximating those of the briefest phonemes in discourse (Watson et al., 1975;

Watson, Kelly, and Wroton, 1976; Watson and Kelly, 1978, 1981; Watson, 1980). Among the variables studied were the abilities to detect frequency, intensity, and durational changes of individual components. The effect of the position of the target item in the sequence was found to be important, with later-occurring sections of the pattern being resolved with much greater accuracy. After long training with minimal uncertainty (concerning the nature of the pattern and the position of the target within the pattern), listeners were able to achieve detection and discrimination for individual components in these sequences that approximated performance for the same components presented in isolation. Interestingly, while under conditions of high uncertainty, target position made a profound difference in performance, and the positional effect was negligible for low uncertainty conditions. Watson and his colleagues found that detection and discrimination of tones dropped to very low levels in pattern contexts when there was a high level of uncertainty, so that it appeared that the sensation levels of tones were effectively very low within the sequences, reaching effective attenuations as great as 40 or 50 dB. In these studies, great individual differences were demonstrated, and in some cases the training time to asymptotic performance was very long—extrapolated to months or even years under some conditions.

Sorkin (1987) employed binary sequences of tones consisting of from 8 to 12 items, having mean tonal durations of 30 to 40 msec and varying intertonal gaps. Listeners were required to judge whether the frequency patterns of two sequences were the same or different. It was found that the temporal pattern had a great effect on a listener's ability to make judgments based on frequency patterns. Sorkin concluded that an extension of the Durlach and Braida (1969) dual model could explain the experimental results. This model considers that two processing modes are available to a listener faced with a discrimination task: a trace mode involving operations performed on a rapidly decaying echoic replica of the stimulus, and a context mode in which operations involve encoded, categorical transforms of long-term stability.

Extended Binary Sequences

Garner and his colleagues (Garner and Gottwald, 1967, 1968; Preusser, 1972; Royer and Garner, 1970) used extended sequences consisting of patterns constructed from two elements (for example, high tone and low tone). They came to three main conclusions: (1) a recognition task gave different results than an identification task; (2) the type of perceptual organization used by subjects changed with the duration of the items; and (3) some sequences were perceived as holistic patterns, without direct identification of the component items.

Repeated or Looped Sequences

The first studies to use repeated sequences reported a surprising inability of listeners to identify the order of components (Warren, 1968; Warren et al., 1969). Listeners heard a sequence of four sounds consisting of successive steady statements of a hiss, a tone, a buzz, and the speech sound "ee." Each item lasted 200 msec, and the sounds were played over and over in the same order. Listeners could not name the temporal arrangement even though the duration was well above the classical limit for detection of order. Although it was possible to identify each of the sounds, the order remained frustratingly elusive. Subsequent experiments used sequences of single types of sounds.

Sequences of tones, of course, are related to music and have a special importance for that

reason. But in addition, they provide stimuli for which it is possible to control perceptual distance between items by adjustment of frequency differences.

It is known that in music, a pair of interleaved melodies splits into two continua when each is played in a separate register. The fact that melodies can be heard under these conditions has been attributed to the dominance of frequency contiguity over temporal contiguity (Ortmann, 1926). The technique of interleaving melodic lines was used extensively by Baroque composers such as Bach and Telemann, so that a single instrument can seem to produce two simultaneous melodies. Such segregation has been called "implied polyphony" by Bukofzer (1947) and "compound melodic line" by Piston (1947). Dowling (1973) has referred to this segregation as "melodic fission." This splitting has been studied in the laboratory by Bregman and Campbell (1971), who used looped sequences of six tones, three in a high register and three in a low register. They found that it was easier to identify the order of tones within as opposed to across these groupings, and they called this segregation "perceptual auditory stream segregation." Thomas and Fitzgibbons (1971) found that successive tones within looped sequences of four items have to be within one-half octave for identification of order at the limiting value of 125 msec per item. However, a decrease in accuracy in the naming of order with an increasing frequency separation was not observed in later studies involving looped sequences of four tones by Nickerson and Freeman (1974) and by Warren and Byrnes (1975). While there is little doubt that a splitting into auditory streams occurs in Baroque compositions, there is a puzzling difficulty in obtaining reliable analogous splitting with nonmelodic looped tonal sequences.

There is another type of perceptual splitting that has been neglected and might be used to further understanding of perceptual auditory stream segregation. Heise and Miller (1951) reported that when a sequence consisting of several tones, each with 125 msec duration, had a single member that differed greatly in frequency from other components, it would appear to "pop out" from the ordered group so that a listener could not distinguish which sounds were preceding and which following. This type of segregation does not involve competing streams, but rather a single stream. Requirements for inclusion of single items in that stream could be investigated.

Looped sequences of speech sounds have been examined fairly extensively. Thomas et al. (1970) and Thomas, Cetti, and Chase (1971) found that the threshold for identifying the order of four concatenated steady-state vowels presented in looped mode was 125 msec. Interestingly, the threshold dropped to 100 msec when brief silent intervals were inserted between the steady-state vowels (perhaps the silence avoided the abrupt transitions from one speech sound to the next, a sound that would be impossible in natural speech). Cole and Scott (1973) and Dorman, Cutting, and Raphael (1975) reported that the addition of normal articulatory transitions between successive items facilitated identification of order with looped sequences of phonemes. Cullinan et al. (1977) employed a number of vowels and consonant-vowel syllables and concluded that lower thresholds for the naming of order were associated with a greater resemblance of the sequences to those occurring in normal speech.

While the order of looped sequences cannot be identified for tonal items below about 125 msec and vowel sequences below 100 msec, there is evidence that permuted orders can be discriminated at much briefer durations. It was reported (Warren, 1974; Warren and Ackroff, 1976) that listeners could discriminate between different orders of three or four items, each having durations as brief as 5 msec for both looped sequences and one-shot sequences. At these short durations, sequences could be differentiated only on the basis of qualitative cues. However, listeners could be taught readily to identify and to name the order of items within these sequences (the ease of learning to name orders indicates

that caution must be exercised to prevent inadvertent training effects in order-identification experiments).

What sets the limit for identification of order in looped sequences? It was suggested by Warren (1974) that the time required for attaching verbal labels to sounds determines the lower limit for direct order identification. This limit is about 125 msec for tones and 100 msec for vowels. The lower value for vowels was attributed to the greater speed of verbal encoding for stimuli in which the sound is the same as the name. Teranishi (1977), working with Japanese vowels, independently made the same observations and came to the same conclusion—that is, verbal encoding determines the limit for direct identification of order within extended sequences.

Unfortunately, it is difficult to compare results obtained with the different experimental paradigms that have been developed for studying perception of sequences. It would be of considerable value to theory development if these parallel lines of investigation could be linked. One such linkage, for example, could be accomplished by combining the use of 10-item word length tonal sequences (which are conventionally employed as single statements) with the procedure of looping or repetition (which usually has been used with three- or four-item sequences).

Relation of Repeated Sequences to Other Types of Periodic Patterns

Repeated sequences can be considered special types of periodic sounds in which the iterated patterns consist of discrete elements. We have seen that a holistic perception of the patterns or "temporal compound recognition" operates when the sequence items are brief (below 100 msec) so that the iterated patterns of four items last 400 msec or less. However, there are repeated patterns *without discrete* elements, and the question arises whether similar rules govern pattern recognition for both types of stimuli. The answer to this question would be of interest to theory development.

Perception of Random Patterns Without Discrete Elements

The discovery by Guttman and Julesz (1963) that the iteration of frozen noise segments can be readily detected at frequencies from about 0.5 through 20 Hz has led to a number of studies dealing with different aspects of the perception of long-duration random patterns. In their pioneering work, Guttman and Julesz described the sound of iterated frozen noise as "whooshing" from 1 through 4 Hz, and as "motorboating" from 4 Hz through 20 Hz. While detection of repetition at 1 Hz and above was described as "effortless," repetition could be detected with some difficulty down to 0.5 Hz by skilled listeners. Subsequently, investigators have been interested in the use of repeated frozen noise as a measure of short-term memory (see discussion of "echoic storage" by Neisser, 1967; and "tape-recorder memory" by Norman, 1967). Cowan (1984) has reviewed the use of repeated frozen noise to study such "short-term auditory stores." Pollack has reported a number of studies dealing with the nature of information processing involving repeated frozen noise. Initially, he had a hunch that listeners perceived repetition through detection of the iteration of extreme amplitude components, a mechanism that would require a minimal storage involving only unusual events. However, he found that when frozen noise segments were modified so that the normal distribution of amplitudes was spaced at only +5 percent about the mean, repetition was still detected with ease (Pollack, 1969). In subsequent publications, he has explored the rules governing the perception of repeated frozen noise through modification of the temporal microstructure of the stimulus in various ways (Pollack, 1975a, 1976a, 1976b,

1978, 1983). He also has investigated the ability to retain the memory of long-duration random patterns and to recognize them when presented with a pool of alternative patterns with similar durations and long-term spectral characteristics (Pollack, 1972, 1975a), as did Pfafflin and Mathews (1966) and Schubert and West (1969). While these studies found that listeners could remember the patterns and identify them when presented with an array of similar patterns, the task was far from easy and was accompanied by considerable uncertainty and confusion during the early stages of training.

A recent study attempted to relate perception of iterated frozen noise segments to repetition of sequences of discrete sounds (Brubaker and Warren, 1987). Frozen noise segments were divided into three sections of equal duration (A, B, C), which were reassembled and arranged to form two periodic sounds (ABC)n and (ACB)n. Discrimination between orders was accomplished readily when the duration of A + B + C was 300 msec or less, indicating that a holistic recognition of patterns took place rather than detection of the repetition of singularities for repetition frequencies in the "motorboating" range.

There is some evidence that similar rules govern the perception of acoustic repetition in the pitch and infrapitch ranges. Helmholtz noted that there were two types of listening strategies that could be adopted with complex tones: a synthetic mode that resulted in perception of a fused auditory image with an ensemble pitch corresponding to that of the spectral fundamental; and an analytic mode, in which individual harmonic components could be teased apart. In an attempt to determine whether similar modes existed for harmonically related waveforms with infrapitch periodicities, Warren and Bashford (1981) mixed pairs of iterated frozen broad band noises in the "motorboating" and "whooshing" ranges that had frequency ratios of 1:2, 2:3, and 3:4. They found that while the ensemble periodicity (or waveform repetition rate) with the relative frequency of unity was dominant for each of the three ratios used, listeners could also hear each of the harmonically related repetition frequencies of the mixture. It would be of interest to determine if other phenomena observed for the pitch range of acoustic repetition have analogs at long-period infrapitch repetition rates.

GENERAL PRINCIPLES OF PERCEPTUAL ORGANIZATION

The literature offers only a few hints of general principles for perceptual organization. The Gibsonian view argues that perceptual classification is based as much on knowledge about the objects that generate the sound as on the sound itself. The work of Bregman and his colleagues on stream segregation is largely an attempt to describe properties of sound that may form figure (foreground) and ground (background) in a complex sound field. Both the ecological approach of the Gibsonians and the hypotheses concerning the formation of auditory streams have their foundation in Gestalt principles.

One way to classify an object is to identify it. A hand clap is classified as a hand clap, rather than by its perceptual attributes (e.g., its timbre) or its acoustics (e.g., its attack time). It is the actual object or event, not some transform, that determines its perceptual classification. This source or event perception represents a weak form of ecological perception (Gibson, 1976; Neisser, 1976), which is discussed Chapter 6. This approach to perception has been used in vision to some extent, but almost not at all in hearing outside the areas of speech perception and music. Knowledge about the source of the object or event may also serve as a means of classification. One form of this approach can be found in the so-called motor theory of speech perception (Liberman and Mattingly, 1985). That is, the perception of the parts of a speech sound is derived from knowledge about how that sound is produced. A particular consonant, for instance, is perceived as such because that

sound can only be made in a unique manner by the speech mechanism. The nervous system classifies the sound as that consonant because it "knows" how the consonant was produced. In music, an obvious use of the ecological approach is to classify music sounds by the instrument generating the sound. However, there is no general auditory theory concerning an ecological approach to classifying sounds outside music and speech.

A few recent studies (Jenkins, 1985; Repp, 1987; Freed, 1987; Warren, 1986) have investigated complex sound (nonspeech and nonmusic) perception and classification in the context of dealing with the source. Repp's study (1987) of hand clap perception can serve as an example. The basic questions are: What acoustic and/or perceptual variables distinguish one hand clap from another, and which variables are most salient for classifying a particular hand clap (say the ability of one person to correctly identify his or her own hand clap)? A physical spectral or temporal domain measurement of a variety of hand claps is made. Then the physical measurements are subjected to a factor analysis, discriminant feature analysis, or multidimensional scaling analysis (see Kruskal, 1964a, 1964b; Shepard, 1982) to determine which physical variables account for most of the variance in the differences among the hand claps. Human listeners can then be asked to judge the similarities among the same hand claps. Correspondence between the human judgments and the physical measurements may be used to infer the bases for perceptual classification of hand claps. A more detailed perceptual study may involve modification of the recorded hand clap signals. The physical properties of the hand claps can be altered (along the lines suggested by the multidimensional analysis described above), and the listener's ability or inability to judge the hand claps can be investigated as these physical variables are altered. For instance, if attack time appears to be an important physical variable for accounting for the variability in the measure of the hand claps, then the hand claps can be altered to have only onsets or only offsets. Presumably, the onsets would allow for judgments similar to those measured with the entire waveform, while offsets would not.

The technique described above has proven valuable in both speech and music (timbre) perception (see Repp, 1987, and McAdams, 1984b, for reviews). It is possible that a series of such studies for nonmusic and nonspeech sounds might reveal some basic properties that listeners use to classify ecologically relevant sounds. At the moment this work involves brief sounds and suggests that the spectral characteristics of the sound's attack are the most important physical parameters governing this weak form of ecological perception.

5
Limits of Auditory Processing of Complex Sounds

ROLE OF MEMORY

In order to classify sounds into disjoint groups (e.g., according to source) it is sufficient, though perhaps not necessary, to be able to identify the sounds. During the past 30 years our understanding of the factors that limit the accuracy of sound identification has increased significantly, particularly for simple sounds. Sound identification has generally been studied in experiments in which listeners attempt to identify sounds in accordance with an objective payoff function (i.e., there is an experimenter-defined correct response for each stimulus). In addition, certain relevant studies have employed roving-level discrimination experiments (in which, for example, the listener must judge whether the second of two sounds is louder or softer than the first, while the overall level of the pair varies randomly over a range of levels), sorting (binary sorting with possibly irrelevant variation of one or more physical attributes of the stimuli), and similarity scaling.

The classic studies of Pollack (1952) and Garner and Hake (1951) demonstrated conclusively that our ability to discriminate properties of sounds (e.g., loudness, pitch) generally exceeds by a substantial factor our ability to identify the values of those properties absolutely. Identification accuracy is constrained to a channel capacity of only 2-3 bits for such properties by most listeners (a significant exception being the phenomenon of absolute pitch), corresponding to accurate categorization with 4-8 categories. When several properties of sounds are varied independently, the number of sounds that can be identified increases, but there is loss of accuracy for each component property (Pollack and Ficks, 1954). These phenomena appear to be quite general: similar results have been obtained in the visual, tactual, and gustatory senses. A substantial portion of the research on categorization since these studies has focused on two problems: understanding the factors that limit the ability to identify a given stimulus attribute (unidimensional categorization) and understanding the interactions between attributes (multidimensional categorization).

Research on unidimensional categorization has focused on the effects of varying the number and range of sounds to be identified, the distribution of sounds within a given range, payoffs, the a priori presentation probabilities, and the availability of reference sounds. Much of this work has been conducted on the identification of sound intensity by Durlach and Braida (1969) and their colleagues at the Massachusetts Institute of Technology. Results have generally been reported in terms of the sensitivity measure used in the theory

of signal detection (e.g., Green and Swets, 1974), d', rather than percentage correct or the information transfer measure (mutual information) used in the classical studies. Many results are conveniently summarized in terms of total sensitivity, the sum of the d's between adjacent stimuli. (When stimuli are uniformly distributed throughout a given range, the most common experimental condition, mutual information grows roughly logarithmically with total sensitivity, provided total sensitivity is greater than about 2.) Total sensitivity was found to be relatively unaffected by changes in the number of sounds to be identified within a given range (provided the number is five or more), in contrast to what might be expected if identification accuracy were limited by an inability to remember fixed sound prototypes. However, total sensitivity was found to depend on the intensity range of the sounds, being proportional to range for small ranges (e.g., Braida and Durlach, 1972; Pynn, Braida, and Durlach, 1972) and reached an asymptote at a constant value for large ranges. In two-interval roving-level discrimination experiments, sensitivity to a given stimulus increment was found to decrease as the interstimulus interval increased, but at a rate that depended on the range of overall level variation (e.g., Berliner and Durlach, 1973a; Berliner, Durlach, and Braida, 1977), with greater decreases observed when the range was large. When the interstimulus interval was long, sensitivity in the discrimination task was comparable to that found in an identification task with the same range of intensities.

Within a given range, the ability to resolve two intensities in an identification experiment was found to be roughly independent of the distribution of intensities. Neither moderate changes in presentation probabilities (Chase et al., 1983) nor moderate changes in payoffs (Lippmann, Braida, and Durlach, 1976) were found to have significant effects on sensitivity, provided the listener was expected to attend to the entire range of intensities. More extreme variation of presentation probability was found to cause some improvement in sensitivity in the vicinity of the most frequently presented stimuli, as was an extreme simplification of the judgmental task (Nosofsky, 1983a). The relative invariance of sensitivity to changes in a priori probabilities or payoffs contrasts with marked changes in response bias, which are in the appropriate direction although smaller in size than would be expected for optimum performance. The relative constancy of sensitivity when a priori probabilities or payoffs are varied presumably reflects a listener's inability to focus on a subrange of intensities when intensities outside the subrange must also be identified.

The availability of stable perceptual references (e.g., explicitly presented standards) would be expected to improve information transfer to the extent that they permit the listener to bifurcate the stimulus range unambiguously (e.g., Pollack, 1953). Berliner, Durlach, and Braida (1978) found that an explicitly presented standard intensity increased sensitivity in the region of the standard when the range was large, provided the standard corresponded to a mid-range intensity rather than to an extreme intensity. When the range is small, the availability of a standard has, by comparison, a smaller effect on sensitivity (e.g., Long, 1973). Durlach and Braida (1969) and Braida et al. (1984) have interpreted the results of these studies as reflecting the effects of two types of limitations on performance: those associated with imperfect sensory mechanisms (which are presumably independent of the experiment), and those associated with imperfect memory mechanisms. They assume the existence of two memory mechanisms: a trace-maintenance mechanism (e.g., Kinchla and Smyzer, 1967) whose accuracy decreases with the passage of time, but at a rate independent of range, and a context-coding mechanism whose accuracy is inversely proportional to range but unaffected by the passage of time. Performance in identification experiments is generally limited only by sensory factors and the context-coding mechanism.

According to the perceptual anchor model of context coding (Braida et al., 1984), intensities are identified by estimating the locations of the sensations corresponding to

the intensities relative to well-maintained perceptual references but using an inaccurate measuring process. By assuming that sensations and anchors are corrupted by roughly equal additive noise, and that the measurement of distances is accomplished by counting steps using a noisy ruler (which divides the distance between anchors into a fixed number of steps, independent of range), Braida et al. (1984) were able to account for both the dependence of sensitivity on range (the range effect) and the variation of sensitivity within a given range (the edge effect). The edge effect, found in both one-interval and two-interval experiments when the range is large, increases relative discriminability for stimuli near the edges of the range, corresponding to the putative locations of the perceptual anchors. Additional evidence for the existence of perceptual anchors at the extremes of the range comes from the failure of explicit standards to improve performance when presented at the extremes of the range (e.g., Berliner et al., 1977). Berliner et al. (1973b) and Marley and Cook (1984) have developed alternate forms of the anchor coding model that lead to similar predictions for the edge effect and the range effect.

Relatively little is known about the processes that determine the locations and variability of the anchors used in absolute judgment. It seems likely that overall performance would be improved if stable anchors could be maintained within the stimulus range. Listeners must be able to adjust anchor locations to bracket the stimulus range to achieve the improvements in sensitivity associated with the edge effect. In large-range magnitude estimation experiments, sensitivity is lower than in absolute identification and more uniform throughout the range, as one would expect if listeners were using anchors spaced considerably away from the edges of the range, toward the natural extremes of the dynamic range of the continuum judged. Luce et al. (1982) have shown that when stimuli are not selected uniformly within the range independently from trial to trial, but rather satisfy severe sequential constraints (only $\pm$ 5 dB changes from trial to trial), sensitivity improves and the relative size of the edge effect decreases. This improvement would be expected if listeners could dynamically adjust anchor locations to bracket the region of intensities highly probable on a given trial.

An alternative account of the range effect has been provided by Gravetter and Lockhead (1973), who assume that criterion range rather than stimulus range determines the accuracy of absolute judgments. In identification experiments with uniform stimulus spacing, this account makes predictions roughly equivalent to those of the perceptual anchor model, but for distributions of stimuli clustered in the middle of the range with only a few extreme intensities, it predicts increased sensitivity relative to uniform spacing. While some improvements consistent with the model have been observed, both Nosofsky (1983a) and Green (1988) have found that increasing the stimulus range decreases sensitivity in tasks that require only a binary response (and presumably a single response criterion).

An alternative account of the range and edge effects is provided by the dual-process attention band model (Luce, Green, and Weber, 1976), which assumes that while coarse discriminations can be made over the entire stimulus range, fine discriminations can be made only within a narrow (roughly 10 dB) attention band. However, Kornbrot (1980) has shown that this model is not capable of predicting in detail the confusion matrices typically observed in identification experiments. When more than one of the distinct perceptual properties of the sounds vary, the ability to identify the sounds can improve. The extent of the improvement depends on the performance that is achieved on each property separately and the nature of the covariation of the properties present in the stimulus set. Studies of identification performance under such conditions have generally been less systematic than for the case of a single perceptual property. For example, there has been little study of the effect of varying the range and number of values for each distinct property. In addition, subjects have generally been less extensively trained in the identification task. Although

most of the studies reviewed have employed visual rather than auditory stimuli, the more salient of these studies were reviewed because it seems likely that many of the factors that determine the speed and accuracy with which complex auditory displays can be categorized would be related to those that limit usual categorization.

When identification performance is characterized in terms of information transfer, the largest improvements have been observed when the stimulus set is derived by varying independently several distinct physical dimensions of the stimulus. For example, Pollack and Ficks (1954) found that listeners could achieve 7.2 bits information transfer (equivalent to roughly 150 categories) per stimulus presentation when six acoustic variables (intensity, frequency, interruption rate, duty cycle, duration, and spatial location) were each allowed to assume one of five possible values. Several aspects of these results are of interest. First, when more than one physical property must be judged to identify the stimulus, the accuracy that can be achieved in identifying each property is less than when only one property must be judged. As a result, the information transfer that can be achieved when two or more properties must be identified is less than the sum of the transfers that can be achieved for each property separately. Even when the properties assume only perfectly discriminable binary values, errors are made when simultaneous identification of several properties is required. In addition, the time required to perform such identifications increases, so that the information transfer rate does not necessarily improve.

Egeth and Pachella (1969) have argued that the decreased ability to identify values of each of the component stimulus properties stems from four factors: reduced observation time, differences in discriminability of one property at various values of the second property, distraction associated with irrelevant variation of the second stimulus property, and response complexity. When observers identify the horizontal and vertical coordinates of a visual target, accuracy on a given coordinate is unaffected by irrelevant variation of the second coordinate when this variation need not be responded to, but is reduced when both coordinates must be identified on a given trial. The latter reduction is dependent on the observation interval: decreasing from 0.47 bits/coordinate at 2 sec to 0.09 bits/coordinate at 10 sec. They also found that the accuracy with which a given coordinate was identified depended on whether it was responded to first or second on a given trial, with the first response more accurate by roughly 0.15 bits/coordinate independent of the observation interval.

A second way to improve information transfer is to construct a stimulus set in which several properties of the stimulus covary in a regular fashion. For example, in a visual task, Eriksen and Hake (1955) found that squares that covaried in size, hue, and brightness could be identified more accurately than squares that differed in size, hue, or brightness alone. The average gain in information transmission was 0.43 bits when two properties covaried and 1.03 bits when three properties covaried. Garner and Creelman (1964) obtained similar increases when hue and size were covaried at presentation durations of 0.04 and 0.10 sec. Lockhead (1966) found that horizontal line segments that varied in both length and vertical position could be identified more accurately than similar segments that differed only in length or position (average gain 0.13 bits/presentation).

Lockhead (1970) found the increase in information transfer for correlated properties to depend markedly on the nature of the correlation between properties. For example, when the hue and brightness of colored patches were covaried in a uniform fashion (so that small changes of hue corresponded to small changes in brightness), information transfer was 2.0 bits per stimulus. However when the covariation was "sawtooth" in nature, an additional 0.5 bits were transferred. Similar advantages for the sawtooth condition were observed for complex stimuli consisting of pairings of visual brightness with auditory loudness and

of visual hue with tactile roughness. He also found that information transfer continued to increase when additional properties were covaried in a sawtooth fashion. For example information transfer for the identification of position and hue of colored discs presented on backgrounds of varying brightness (10 total stimuli) was 2.75 bits, compared with roughly 1.2 bits for each property alone. Even more dramatic increases were observed when the stimuli were constructed by varying the angular orientation of four line segments in a correlated fashion: errorless identification of 20 stimuli (4.3 bits) was achieved compared to 32 and 37 percent correct identification for single segments and for uniform correlation, respectively.

Nosofsky (1983b) showed that sensitivity in intensity identification could be improved by using multiple stimulus presentations. For example, with three observations, sensitivity increased by roughly 40 percent relative to a single observation, for both narrow (10 dB) and wide range (32 dB) conditions. Since increases in stimulus duration do not generally improve identification performance (e.g., Garner and Creelman, 1964), it appears that the gain in accuracy results from an improvement in the coding process rather than from reduced variability of the sensory representation of the stimulus. In this sense, multiple observations of a given intensity should produce roughly the same improvement in accuracy as uniform covariation of two stimulus properties.

Garner (1970) argued that the ability to classify stimulus sets that contain variations in several stimulus properties depends on the perceptual relation between the properties. Integral properties are described by a Euclidean metric in similarity scaling while separable properties are described by a city-block metric. Integral properties cannot be perceived selectively, so that when correlated variation is introduced, accuracy and speed of classification should increase; when orthogonal variation is introduced, classification should be degraded. For separable properties, neither correlated nor orthogonal variation should affect classification. In studies of speeded classification of colored chips, Garner and Fefoldy (1970) found saturation and brightness to be integral when variations were combined in one chip, but separable when variations were presented in two chips. However, the pattern of performance when both properties were varied in a single chip was found to depend on relative ranges of the two properties: there was less interference in the orthogonal condition when the range (or discriminability) of one of the two (binary valued) properties was increased.

Nosofsky (1985) analyzed a visual identification experiment in which the stimuli were semicircles varying in size and in the orientation of an interior radial line. Although these properties were expected to be separable (e.g., Shepard, 1962), the Euclidean metric with a Gaussian similarity function described the similarity properties of the confusion matrices better than a city-block metric. In analyzing the discrepancy with Shepard's findings, Nosofsky argued that the resolution edge effect (well established in intensity identification) would be expected to distort predictions based on a Euclidean metric toward those based on a city-block metric. Shepard (1982) has suggested that the similarity structure expected for separable stimulus properties may depend on overall discriminability.

Several trends evident in recent studies of the classification of complex visual patterns seem likely to have relevance for the auditory classification of complex sounds. When several properties of the stimuli to be classified vary, speed and accuracy of classification are affected differently for separable and integral properties. Consistent descriptions of stimulus properties as integral or separable depend on the convergence of operational measures, including similarity scaling. Detailed mathematical models of complex stimulus classification are only beginning to emerge, and the specification of factors that affect the structure of the perceptual space in the complex classification task is at best incomplete. Some increase in clarity

seems likely to result from incorporating the improved understanding of the factors that determine identification performance for simple sounds in studies of complex classification.

UNCERTAINTY AND ATTENTION

In general, uncertainty about the spectral or temporal structure of a sound interferes with the ability of a listener to extract information from, or about, the sound and its source. Early studies of uncertainty effects with simple sounds demonstrated small but consistent reductions in performance when some aspect of the sound or its presentation was uncertain. An illustration of this point, and perhaps the first detailed theoretical consideration of uncertainty in hearing, is found in studies comparing the detectability of a pure tone of random or uncertain frequency with a tone of known frequency. This literature dates back at least to the report by Tanner and Norman (1954) and is summarized in Green and Swets (1974). Relative decreases in discrimination performance have likewise been reported in uncertain frequency discrimination tasks (e.g., Harris, 1952; Jesteadt and Bilger, 1974), uncertain intensity discrimination tasks (e.g., Berliner and Durlach, 1973a), and tasks requiring the detection of sounds occurring at uncertain times (e.g., Egan, Greenberg, and Schulman, 1961; Green and Weber, 1980). While the studies of the effects of stimulus uncertainty on the detection or discrimination of simple acoustic signals do not directly address the classification of complex sounds, they do raise important issues concerning mechanisms that are likely to be relevant. Some of these issues include: the fine-tuning of sensory mechanisms due to attention (e.g., Sorkin, Pastore, and Gillom, 1968; Luce et al. 1976; Swets, 1984), sequential effects (e.g., Purks et al., 1980; Luce et al., 1982), and perceptual anchors (e.g., Braida et al., 1984; Macmillan, 1983).

The role of stimulus uncertainty in the perception of sound patterns has been studied extensively by Watson and his colleagues. A typical paradigm they have employed is one in which the listener must detect an alteration in the pattern formed by a series of sequentially presented tones. The nature of the "alteration" is often a difference in the intensity, frequency, or duration of a single component of the pattern, allowing comparison with much of the traditional research on discrimination. Watson and Kelly (1981) provide a review of a portion of that work. They describe effects as large as 40 to 50 dB, comparing some highly uncertain stimulus conditions with minimally uncertain stimulus conditions attributing many of the effects to "informational masking" (Pollack, 1975b).

Recently, studies reporting the effects of spectral uncertainty in the perception of complex sounds. The series of papers on auditory profile analysis began with attempts to quantify and explain the relatively small effects of spectral uncertainty on the detection of spectral shape alterations (e.g., Spiegel, Picardi, and Green, 1981). Much larger effects of spectral uncertainty have been reported by Kidd, Mason, and Green (1986) and Neff and Green (1987) for conditions in which a different spectral pattern was present for every stimulus during the procedure.

To the extent that the perception of differences among tonal patterns or among spectral shapes involves the assignment of different stimuli to signallike or nonsignallike categories, these studies may provide the best indications to date of the effects of uncertainty on the classification of complex sounds. In those experiments, the random composition of the stimulus ensemble requires that the listeners group sounds according to similarity along one or more stimulus dimensions, while ignoring irrelevant stimulus differences within groups. Clearly, uncertainty interferes with that process and may limit performance. Within that context, although the above studies of uncertainty in complex sound perception are certainly relevant to the topic, they were not designed to study classification per se.

Future research in this area, therefore, could address a wide variety of issues. Could the effects of uncertainty be modelled simply by assuming noise is added to the stimulus or to the sensory transduction process? Or is uncertainty better considered with respect to the expectations about the stimulus properties the observer may have formed prior to presentation of the sound or sound sequence? Watson and Kelly (1981) consider the possibility that the attentional control of the sensory mechanism degrades the stimulus representation from the auditory periphery in highly uncertain conditions. Thus, a case could be made for uncertainty having both peripheral and central components. The issues of bottom-up versus top-down processing in the classification of acoustic patterns have also been pointed out by Howard and Ballas (1980). Their finding that prior semantic knowledge about a sound may, in some instances, interfere with extracting information about pattern structure argues for at least some top-down processing and allows the interpretation of uncertainty effects to include misinformation. In general, however, there are too few studies to accurately predict the effects of various forms of stimulus uncertainty on classification of complex sounds. More research is needed in order to develop broadly based models of uncertainty and to evaluate adequately the applicability of current models of audition to uncertainty effects.

LIMITATIONS DUE TO INTERNAL NOISE

Our ability to detect auditory signals appears to be limited by the presence of internal perturbations or noise. Many studies have attempted to characterize the magnitude and character of the internal noise. Swets et al. (1959) required subjects to make repeated observations of identical or independent noise samples in order to determine the relative improvement in detection performance as a function of the number of observations. Other experiments have assessed the consistency of an observer's performance on identical noise trials (Green, 1964), examined the discrimination of Rayleigh noise (Ronken, 1969) or reproducible noise (Raab and Goldberg, 1975), or examined the difference between observer performance on trials when the noise samples were identical and different (Siegel, 1979; Spiegel and Green, 1981). Virtually all these studies have concluded that the magnitude of the internal noise depends on the magnitude of the external noise. Estimates for the ratio of internal to external noise have varied from approximately 0.3 to 3, depending on the particular task.

The relatively large magnitude of internal noise implied by most of these experiments has led some investigators to study how an observer's response depends on particular attributes of the stimulus input (Pfafflin and Mathews, 1966; Pfafflin, 1968; Ahumada and Lovell, 1971; Ahumada, Marken, and Sandusky, 1975; Hanna and Robinson, 1985; Gilkey, Robinson, and Hanna, 1985; Gilkey and Robinson, 1986; Gilkey, 1987). In these studies, performance is observed over a number of repeated presentations of the same stimulus. The results seem to indicate that an observer employs a more complex observational strategy than previously thought. That is, an observer probably does not use a single, fixed-bandwidth filter located at the signal frequency, or a single integration window matched to the signal's occurrence. Instead, the observer may compare information obtained from several different spectral regions and at different times. Some of the apparent variability of an observer's responses may be due to the use of information outside the immediate temporal and spectral region of the signal. The idea that an observer's decision may be based on the weighted combination of energy sampled from different spectral-temporal portions of the input is consistent with the profile analysis hypothesis described in studies by David Green and his

colleagues (Spiegel et al. 1981; Green, Kidd, and Picardi, 1983; Green, Mason, and Kidd, 1984).

Some specific sources of internal variability have been identified. These include noise due to observer uncertainty about signal or noise parameters, noise associated with the processes of encoding or storing the stimuli, and noise associated with the setting and maintenance of response criteria. Tanner (1961) proposed a framework for categorizing the memory requirements of several general types of detection and discrimination tasks. He characterized two-interval (2AFC) performance as being limited by an internal sensory noise plus a memory noise that increased as a function of the time separation between the two observation intervals. This time-dependent memory noise is equivalent to a memory trace that decays over time. Tanner's model has been extended by Sorkin (1962) in a study of the same-different discrimination task, and Macmillan, Kaplan, and Creelman (1977) in more complex discrimination tasks. The trace decay component was expanded by Kinchla and Smyzer (1967) and incorporated as the trace component of the two-component trace-context model developed by Durlach and Braida (1969).

In the Durlach and Braida (1969) theory of discrimination there are two sources of internal variability in addition to the traditional sensory noise: (1) a context noise associated with encoding/identifying a given stimulus from an ensemble of possible stimuli and (2) a trace noise associated with maintaining an accurate representation of a presented stimulus. The context noise is assumed to increase with the total range and number of possible stimuli and is independent of the time held, while the trace noise is assumed to increase with the storage time.

The properties of the trace noise component have been studied in experiments by Berliner and Durlach (1973a) and by Lim et al. (1977), in the context of Durlach and Braida's two-mode theory. Lim et al. studied loudness matching and intensity discrimination for signals of different frequency. An observer's ability to discriminate the intensities of two signals decreased as a function of the frequency difference between the signals; Lim suggested that this was due to a transformation noise component of the trace noise process. Hanna (1984) also studied the limitations on discrimination caused by internal noise of different types, such as sensory variability, memory (trace or context mode variability), and attentional factors and decision-making components (informational masking). He examined the discrimination of reproducible noise as a function of the bandwidth and duration of the noise bursts, the time interval between the bursts, and the effects of forward and backward maskers.

Sorkin (1987) and Sorkin and Snow (1987) applied the Durlach and Braida theory to the discrimination of tonal sequences. They studied the characteristics of trace and context noise in tasks that required observers to discriminate between tonal sequences having the same or different frequency patterns. Trace noise was observed to increase rapidly with the introduction of variation in the temporal structure of the sequences but was relatively insensitive to other envelope decorrelating operations such as uniform expansion or compression of the tonal durations and gaps.

Berg and Robinson (1987) also reported on a task involving tonal sequences; subjects were presented with sequences of tones sampled from one of two probability density functions on frequency; on each trial the subjects had to decide which distribution produced the sampled tones. In their model, internal noise is composed of peripheral variance (noise added to each tone observation prior to formulation of the decision statistic) plus a central variance (noise added to the decision statistic component). Increasing the variance of the probability distributions (while controlling the difference between the distribution means)

resulted in increases in the internal noise, suggesting that the internal and external variance are not independent.

A number of investigators have attempted to relate performance in detection and discrimination to more complex tasks such as absolute judgment and magnitude estimation. Tanner (1956) proposed an extension of the signal detection model to the two-signal recognition task, in which interstimulus distances were defined by performance in separate detection and discrimination tasks. Shipley (1965) and Lindner (1968) tested high threshold models of detection using combined detection and recognition tasks. An extension of signal detection theory to the more general recognition-detection task was reported by Green and Birdsall (1978). In such tasks, the observer may be presented with either no signal or one of n signals; the observer must identify which signal, if any, was present. Experiments reported by Green, Weber, and Duncan (1977) provided reasonable support for a theorem relating signal identification to signal detection performance. This approach also has been applied to the identification and detection of visual signals (Swets et al., 1978).

Attempts to explain the variability in an observer's behavior in recognition and magnitude estimation tasks have involved assumptions similar to those in detection and discrimination, such as internal noise in the basic sensory mechanism and variability of the attention band or response criteria. For example, Green and Luce (1974) discussed the results of several magnitude estimation experiments in terms of a timing theory analysis, in which variability in the observer's responses is a consequence of the assumed internal timing mechanism. Their interpretation of the data included the assumption of an observer attention band, whose location on any trial depended on the nature of the signals on the preceding trials.

A number of experiments employing judgment and estimation tasks have been analyzed by Treisman (1984, 1985; Treisman and Faulkner, 1984, 1985; Treisman and Williams, 1984), using a model in which the observer's criteria are determined and maintained as a function of the expected stimulus values and the observer's prior responses. The general model exhibits several of the effects noted by other workers in absolute judgment and magnitude estimation, including a dependence of performance on the number and range of the stimuli, and on the presence of correlations between successive responses and of edge effects.

Finally, Nosofsky (1983b) reported an analysis of absolute judgment for auditory signals varying in intensity, using a multiple observation procedure. The procedure enabled him to estimate the magnitude of both the stimulus noise and the criterion noise as a function of the range of the stimuli to be identified; both appeared to increase as the stimulus range was increased.

LEARNING

Introduction

Most classification of complex sounds probably is based on some level of prior learning or training. It is well documented that early experience can modify the production and of the responsiveness to calls in a wide variety of animals and birds with species-specific calls, although early learning has not yet been demonstrated for frogs or insects (Ehret, 1987). Early linguistic experience for humans is believed to shape significantly, possibly permanently, the nature of speech perception (e.g., Pisoni et al., 1982). Thus it is quite possible that early auditory experience with nonspeech stimuli also plays a significant role in shaping the perceptual organization of the auditory environment for the individual as an adult. Therefore, the development of auditory perceptual organization, and the role played by early experience, is one important field that needs to be investigated.

Since the major organizational abilities and skill important to the perception of speech stimuli and to the perception of visual patterns are believed to have achieved a high degree of development within the first few years of life, the focus of research on perceptual learning for adults is different than for infants. With adults the focus of important research questions concerns effectiveness of training procedures to modify existing perceptual skills, or to create new skills, for the classification of complex nonspeech sounds. Within this general training focus, there are several major research issues that need to be addressed. There is a need to evaluate the effectiveness of different types of training and to determine whether there are individual or age-dependent differences in ability to learn new perceptual skills or to modify existing skills. Each training procedure needs to be evaluated in terms of the generalization of the training across types of stimuli and stimulus situations. The nature and appropriateness of different perceptual strategies to various types of tasks need to be addressed.

The ability to classify sounds requires some operating knowledge about the important feature(s) by which specific sounds should be grouped together or about the rules, or the means of generating rules, for grouping sounds in an orderly way. By far the most frequent reference to or use of the term *learning* in the literature on nonspeech sound perception is in the context of acquainting listeners with the requirements of a particular experimental task or with internalizing the value of a stimulus along a particular perceptual dimension to be used as a reference. In contrast, learning to attend to specific aspects of a complex sound or sound sequence, which varies along several dimensions simultaneously, and attempting to assign the stimulus to a particular group, has not been studied extensively. The issues involved in learning in audition are complex and diverse, extending across multidisciplinary boundaries.

Learning Complex Nonspeech and Nonmusic Sounds

Many of the studies most appropriate to the topic of learning to classify complex nonspeech sounds concern, or appear to have been motivated by, the study of human detection and identification of underwater sound sources. Webster and colleagues (e.g., Webster, Carpenter, and Woodhead, 1968a, 1968b) considered learning processes associated with the identification of complexes of harmonically related tones having different spectral structures. Other studies directly applicable to underwater sound identification or to the techniques that could be used to train sonar operators are the papers by Howard and others (e.g., Howard and Silverman, 1976; Howard, 1977; Howard and Ballas, 1982). An interesting theme emerging from these studies is the notion of different processing strategies based on the temporal properties of the sound or sounds to be identified.

The identification of steady-state sounds may involve more bottom-up processes because of the time available to extract critical stimulus features. Transient sounds, by comparison, cannot be analyzed in that manner and may depend to a greater degree on prior knowledge about the structure and likely source of the sound.

One paper specifically designed to measure learning to identify complex nonspeech sounds is that of House et al. (1962). They measured learning functions for identification of stimuli varying along one, or more than one, dimension. They found that learning performance improves as stimulus dimensions are added but that when the test sounds imprecisely resembled previously overlearned sounds (i.e., speech sounds), performance worsened.

Psychophysical Abilities

It is well known that practiced psychophysical subjects usually exhibit different patterns of performance than naive subjects. Most measures of detection and discrimination are more stable and indicate greater sensitivity for experienced subjects than for naive subjects, sometimes even for many new or novel stimulus comparisons. The difference between practiced and naive subjects may reflect heightened sensory abilities in the former. It is more probable that practiced listeners may be better able to attend to, focus on, or perceptually isolate critical components, or patterns of components, within complex stimuli. Practice or experience also may result in subjects ignoring certain characteristics of stimuli. Naive subjects may seem to respond to general stimulus patterns and to ignore many subtle aspects of stimuli. Given this complex characterization of practice effects, it is most likely that the relationship between specific types of practice or prior experience and categorization behavior is both complex and to the nature of the categorization task. This complex relationship is obvious in our brief summary of the existing literature on practice effects for complex sounds.

Psychoacoustic studies include such topics as frequency discrimination in musicians versus nonmusicians (e.g., Spiegel and Watson, 1984), comparison between learning to identify unidimensional sounds along different dimensions (e.g., Houtsma, Durlach, and Horowitz, 1987), the acquisition of category boundaries in labeling speech or speechlike sounds (e.g., Carney, Widin, and Viemeister, 1977; Pastore, 1987a), computer-assisted learning (e.g., Swets et al., 1962; Corcoran et al., 1968), and many others.

Discrimination of Tone Sequences

In a series of experiments, Watson (1987) and his colleagues studied the effects of extended practice on the ability of subjects to discriminate changes in the frequency and intensity of individual components in 10-tone sequences. This research has identified a number of important principles characterizing the limits on discrimination ability as a function of stimulus characteristics and position within a sequence. Although subjects were able to perform very fine discriminations for most components in highly familiar sequences, this ability did not generalize directly to new tone sequences (for reviews, see Watson, 1987; Watson et al., 1976; Spiegel and Watson, 1981; Watson and Foyle, 1985).

Leek and Watson (1984) measured improvements in the detectability of tones embedded in tonal sequences with regular practice over periods spanning several weeks. They found that the amount of informational masking could be reduced by 40 to 50 decibels in some cases, a result they attributed to the long-term acquisition of a reference. The relationship between the time course of learning sounds, the complexity of the sounds, and the experimental task was considered in a paper by Watson (1980). He pointed out that achieving asymptotic performance for experiments employing more complex sounds and complicated tasks often took much longer than for simpler detection or discrimination experiments using isolated tones or noisebands. Kidd, Mason, and Green (1986) found rapid improvement in detecting spectral shape differences during the first few hundred trials of practice of naive listeners with continued, gradual improvement extending over many hundreds of trials. Furthermore, they noted that listeners who had been trained to discriminate a difference in spectral shape for a particular reference sound reached asymptotic performance in learning new reference sounds more rapidly than naive listeners. Neff and Callaghan (1987) report considerable individual differences in learning in random spectrum masking experiments. In experiments in which considerable masking was obtained by maskers with very little energy in the critical band containing a tonal signal, some listeners were apparently able

to find a cue to the presence of the signal after many trials, greatly reducing masking, while the performance of other listeners remained essentially constant. The extent to which differences in attention, motivation, prior experience, etc., affect the rate of learning and asymptotic performance in complex sound perception experiments is not well known.

Categorization of Speech and Music

Research on categorization of speech sounds and musical tones also has identified practice effects. The demonstration of categorical perception, which is characteristic of the perception of stop consonants, requires that the discrimination of stimuli drawn from the same labeling category be typically at chance. However, practice with such a continuum of stimuli results in significantly better than chance discrimination for stimuli within a labeling category (Carney et al., 1977; Samuel, 1977; Kewley-Port, Watson, and Foyle, 1987). The typical explanation of these findings is that practice enables subjects to access the finer acoustic characteristics of the stimuli that are largely unimportant for speech categorization. A number of excellent published reports provide a diverse variety of critical reviews of various factors that may contribute to laboratory measures of performance with speech stimuli and the role played by experience (Strange, 1986; Walley, Pisoni, and Aslin, 1981; Werker and Logan, 1985).

Second Language Acquisition

The perceptual skills or abilities to perceive one's first language probably develop very early in life and then are relatively stable over one's lifetime. Acquisition of a second language by adults thus requires the modification of these existing perceptual skills or the development of new, sometimes incompatible perceptual skills. Therefore, the study of changes in perceptual abilities during second language acquisition offers an excellent opportunity to map the development of new (language-specific) perceptual skills and to investigate correlated changes or modification in the perception of the primary language (Tees and Werker, 1984; Walley et al., 1981; Strange and Dittmann, 1984). There is some evidence that the location of category boundaries for the first language may be altered as the second language is acquired (Flege and Hillenbrand, 1987; Werker and Tees, 1984). In one example of a typical training study, Strange (1972, reported in Strange and Jenkins, 1978) attempted to train English-speaking college students to discriminate voice onset time (VOT) differences that straddled the Thai prevoiced-voiced unaspirated boundary at approximately −20 msec VOT, and found improved performance in the region of the Spanish prevoiced-voiced contrast (−4 msec VOT) and within the voicing category (+15 msec VOT). The interaction of new and related old skills of perceptual categorization probably should be considered in future research on the categorization of complex nonspeech sounds. However, the magnitude or importance of such interactions might not be as great as with speech, whereas second language acquisition typically involves both the perception and the production of the new language (Williams, 1979).

Morse Code Learning

Shepard (1962) analyzed four sets of preexisting data published by other authors, which used the 36 standard International Morse Code signals as stimuli. (Each Morse code signal is composed of up to five tones, called dots and dashes. Each dot has length 1 and each dash length 3, and the separating silences have length 1, relative to an interval that depends

on the speed of presentation.) Wish (1967) collected and analyzed data for a set of 32 sounds similar to three-tone Morse code signals, but each containing two internal silences of length 1 or 3. In all cases, the signals were presented at a sufficiently high speed that it was impossible for naive subjects to explicitly analyze them into their components.

Shepard's chief analysis was based on data from Rothkopf (1957) in which the subjects were explicitly screened to be naive about Morse code, and in which their task was to decide whether two signals were the same or different. Applying multidimensional scaling to these data, Shepard discovered convincing evidence that the two chief perceptual characteristics being used by these subjects were the total length of the signal and the proportion of dots to dashes. Wish, analyzing his own data, confirmed use of the two characteristics and demonstrated two other characteristics, namely, the sound-to-silence ratio and whether the first component is a dot or a dash.

Shepard's second and third analyses were based on identification errors by beginning students learning to read Morse code. Here a memory confusion combined with the perceptual confusion, e.g., the three-dot signal (s) was sometimes identified by the label for the three-dash signal (o) and vice versa. As a result, the two chief dimensions in this case were length of signal and its heterogeneity, wherein an all-dot or all-dash signal is homogeneous and a signal having dots followed by dashes followed by dots is heterogeneous. Shepard's fourth analysis was based on identification errors of more rapid signals by intermediate and advanced subjects. The intermediate subjects showed memory confusion between signals that are time-reflections of each other. The advanced subjects, reading very rapid signals, demonstrated primarily perceptual errors based on mistaking the number of components in a consecutive run of dots or dashes.

Musical Illusions

Deutsch (1982) reports that practice tends to enhance the perception of the octave illusion, which is a type of streaming of alternating, dichotic tone pairs. Perception of such illusions probably involves errors in perceptual grouping or streaming. However, Pastore et al. (1986) report that practice with masking and detection conditions can reduce or eliminate the perception of the octave illusion, even though the subjects had never been exposed to the illusion. This apparent contradiction in findings probably reflects the types of different perceptual strategies described above, with the illusion requiring the (incorrect) perception of stimulus patterns, while detection or discrimination requires analysis of the complex stimuli in terms of critical components.

Perceptual Learning

Pisoni (1971) unsuccessfully attempted to produce categorical perception by training subjects to identify two categories of isolated second formant chirps. More recently, Grunke and Pisoni (1982) found that subjects could learn to consistently assign temporal mirror-image acoustic patterns of CV and VC syllables to arbitrary response categories. Subjects responded to both individual stimulus dimensions and to more general stimulus patterns.

Schwab, Nusbaum, and Pisoni (1985) found that modest amounts of training could significantly improve the recognition of synthetic speech stimuli, which can be considered impoverished, distorted speech. In a subsequent study, Greenspan, Nusbaum, and Pisoni (1986) investigated the effectiveness of different types of training on the perception of synthetic speech produced by rule. Training with isolated words improved only the intelligibility of isolated words, while training with sentences increased the intelligibility of both

isolated words and sentences whether or not they had been used in training. Furthermore, training with the same stimuli each day and training with novel stimuli each day were equally effective. In a second experiment, training with a limited set of repeated sentences did not improve the intelligibility of novel stimuli.

In summary, learning to classify complex nonspeech sounds has not been thoroughly studied. The study of learning in nonspeech sound identification is closely related to studies of uncertainty, and a recurring interest is the extent to which the detrimental effects of stimulus uncertainty may be overcome by internalizing a reference pattern through repeated presentation. The knowledge about this topic could be greatly advanced through further research, particularly with respect to synthesizing and unifying the contributions from a wide variety of investigators in diverse areas of inquiry.

6
Lessons From Speech Perception

POSSIBLE RELEVANCE OF SPEECH RESEARCH

Speech is one class of complex acoustic stimuli that has been studied extensively for many decades. Modern speech scientists now understand most of the important properties of the speech production system and have identified important properties of the physical stimuli that alter the perceived categories of speech. The major focus of modern speech research is the understanding of the perceptual system utilizes the information contained in the acoustic signal to perceive speech. The relevance of the speech research literature to the study of the categorization of nonspeech sounds depends on one's beliefs concerning the nature of the perceptual processes for speech. The more common assumption of researchers is that speech perception is based on some form of higher-order processes that are unique to human speech mechanisms. One alternative formulation of this specialized view is that speech perception is mediated by a separate module that exists at a peripheral level in parallel with other modules specialized for processing acoustic and other specific types of sensory information (Liberman and Mattingly, 1985). If these specialized views are valid in the extreme, then the extensive, and very successful speech perception literature can provide only an example of strategies and techniques for the study of categorization. Although, in principle, our working assumption in the following section is consistent with this majority view of speech perception, there are strong reasons to advise caution in accepting this assumption as valid.

Some researchers at the other extreme argue that speech may not be based on unique, highly specialized processes. The basic approach of these researchers assumes that many of the apparent perceptual differences between speech and other acoustic signals may be artifacts of the largely independent development of the research fields (e.g., Diehl, 1987; Pastore, 1981; Pisoni, 1987; Schouten, 1980). These researchers believe speech perception may be based on higher-order stimulus processing that is largely learned and has developed, at least in part, to make use of unique properties of human auditory signal processing. If this minority view is valid, then much, if not all, of the extensive literature on speech perception may be very relevant to the topic of this report.

There also are various intermediate views on the nature of speech, each with different possible implications for the categorization of auditory sounds. For instance, Stevens argues that the auditory system responds to sounds with different complex acoustic properties in distinctive ways that are important to the classification of sounds that serve as the basis of language (Stevens, 1980). Independent of the validity of the assumed relevance to language,

the efforts to identify the complex acoustic properties are relevant to the classification of all sounds. Other researchers argue that properties of the speech signal are used to convey to the listener static and dynamic characteristics of the vocal tract producing the signal. While the vocal tract is unique to speech sounds, the principles derived from studying the potential and perceived relationships between sound characteristics and source characteristics have very general application. The proceedings of a 1986 NATO conference on the psychophysics of speech perception (Schouten, 1987) provide an excellent summary of the current status of research on speech perception, related aspects of auditory perception, and the relevant psychophysical methodology.

The theory of intensity perception initially proposed by Durlach and Braida (1969) and subsequently developed by these researchers and their coworkers provides a needed conceptual basis for describing and comparing the psychophysical techniques that have been employed in the study of speech categorization. Although the theory is not described here, excellent recent reviews of the theory include a chapter summarizing the general theory (Braida and Durlach, 1986) and a chapter applying the theory to categorization research (Macmillan, Braida, and Goldberg, 1987).

In most of the literature, fixed-level discrimination refers to the condition in which only two stimuli are ever presented in a block of trials. Fixed discrimination represents minimum uncertainty for the given stimulus parameters in the sense that the subject must deal with only the stimulus differences between the two stimuli and the internal system noise associated with that stimulus (trace) coding. Roving-level discrimination refers to the condition in which a number of different stimuli are presented in a block of trials, even though the differences between stimuli compared within a block of trials may be held constant. Roving discrimination represents high uncertainty in that the nature of the stimuli being compared on a given trial is not defined (other than being a member of the broad stimulus set) until the first stimulus is presented. In roving discrimination the subject is faced with the additional variability of the stimuli across trials; in the Durlach-Braida theory, stimulus context coding must be added to stimulus trace coding in performing the task. Sorkin (1987) provides an excellent example of the application of these notions to the perception of complex tonal sequences.

BROAD OVERVIEW

Much of the research on human speech has focused on the relationship of categories of perception to both the acoustic stimuli of speech and the structures of production (or articulation) that normally produce the acoustic stimuli. This study of the relationship between (a) the characteristics of the sound production source, (b) spectral and temporal properties of sound, and (c) categorical properties of perception, represents a type of working structure for future studies of categorization of naturally produced acoustic stimuli (animal calls, engine noises, speech and speaker recognition, etc.), whereas the source properties probably are not important for the categorization of artificially coded cues (e.g., types of alarms, cues for the status of equipment, or even the recoding of information by equipment monitoring aspects of the environment, etc.).

One of the most fundamental questions concerning the categorization of speech stimuli, or any other types of stimuli, is: What aspects of the stimuli cue elicit or give rise to the perception of one category as opposed to another? Specific assumptions concerning the nature of such cues are discussed below in the section on models. The problem addressed here concerns the need for the cues to possess properties that are invariant across a wide variety of source characteristics and listening conditions. This notion of stimulus or cue

invariance need not require absolute or stationary stimulus properties, and it may often require a specification of complex, relative properties. The discussion below is intended to provide a background against which the notion of categorization cue invariance can be better understood.

Natural speech originates from numerous different speakers whose output is highly variable both across speakers and within speakers across time. The listener must perceive an equivalence of stimuli from within a speech category despite sometimes very highly discriminable differences in the physical stimuli. Therefore, if the stimulus attributes that serve as the basis for the perception of speech categories are invariant, that invariance is somehow relative to the context of a high variability. The research literatures on speech coarticulation effects (e.g., Fowler, 1981a, 1981b; Ohman, 1967; Mann and Repp, 1980; Raphael, 1972) and trading relations (e.g., Repp, 1982; Parker, Diehl, and Kluender, 1986) are relevant to this issue. In addition, the research on machine recognition of speech (e.g., Rabiner and Levinson, 1981) and the research on human and machine speaker recognition (e.g., Schmidt-Neilsen and Stern, 1985; O'Shaughnessy, 1986) are quite relevant, since in each type of study one must deal with the extraction of specific categories in the context of highly variable signals (e.g., Garrett and Healy, 1987). This high degree of stimulus variability means listeners are typically operating in a high uncertainty situation.The models of categorization discussed below really differ in terms of whether the assumed critical stimulus properties are general characteristics of the central tendencies of category stimuli or of the boundary between stimuli, and whether the critical category characteristics are properties of the stimuli or of the objects giving rise to the stimuli.

PSYCHOPHYSICAL PROCEDURES

The psychophysical approaches and procedures employed in speech perception studies have tended to be less rigorous than those used in the detection and discrimination studies. While this apparent lack of rigor may seem to represent a problem with this literature, the use of more rigorous psychophysical techniques may represent an analysis that is too fine-grained to study categorization (we return to this point shortly and in the section on categorical perception). The psychophysical techniques employed for speech studies also have not been subjected to the same types of rigorous theoretical and empirical evaluation, although the work of Macmillan (e.g., Macmillan et al., 1977, 1987) represents an excellent beginning.

The important issue concerning psychophysical procedures is not what tasks are best in an absolute sense, but rather what approaches and procedures are most effective and appropriate to studying the relevant question. Pisoni and Luce (1987) and Pastore (1981) provide summaries of different types of psychophysical approaches to the study of speech and simpler acoustic stimuli, as well as the differences in the questions being addressed.

Labeling

Much of the research on speech perception has employed labeling tasks that, almost by definition, are roving-level tasks. In such tasks subjects respond to each stimulus with a label, and categorization is evaluated in terms of the distribution of labels across the stimulus dimensions manipulated. In speech these labels are obvious. In some studies the set of possible labels is limited by the researcher, while in other studies the set of labels is open-ended. Labeling may be viewed as a quick, but imprecise, measure of categorization. Labeling tasks differ from discrimination tasks in several important ways. Labeling results

may be significantly altered by the response tendencies of the subjects. Labeling tasks also are asking a very different question of the subjects: rather than asking if the subject can detect a difference between two stimuli, the subject is asked if the stimuli are sufficiently similar to be categorized as equivalent, as indicated by use of the same label. Therefore, two speech stimuli that are highly discriminable may receive the same label because they are perceived as, or are acceptable as, members of the category designated by one label. One critical consideration in implementing a labeling task is to define a set of response labels that is appropriate for the specific problem or question being investigated. In many labeling tasks the subjects are allowed only a limited set of response labels, or the subjects may conclude that the researcher wants them to use only a limited set of labels, resulting in the assignment of some stimuli to one labeling category rather than to a more appropriate, but unavailable, category. However, when the set of response labels is too open, subjects may indicate perceived differences where categorical equivalence may be more appropriate for the question being investigated.

Some nonspeech stimuli have natural categories for which labels may be (or seem) obvious, such as engine noises, machine shop noises, public address system announcements, etc. Other nonspeech stimuli, especially those initially unfamiliar to the listener, may not have obvious labels. Research on nonspeech and on impoverished analogs to speech-stimuli are excellent examples of research using stimuli for which the response labels must be based on initial exposure to broad sampling stimuli, or to stimuli from the end-points of the stimulus continuum under study (e.g., Pisoni, 1977; Pastore, Harris, and Kaplan, 1982). Finally, there may be a hierarchy of categorization levels. For instance, the category of engine noises can be divided into the subcategories of tuned and malfunctioning engine noises. The engine noise category also can be subdivided into subcategories of noises from motor vehicles, ships, and aircraft, with each divided into more precise categories such as propeller and jet aircraft noises. With considerable accuracy, a highly trained listener may be able to label noise from a single type of aircraft based on the engine manufacturer. With such different levels of categorization, the researcher must be careful that the subject is operating at the expected level, and that comparison of results across laboratories, or even within a laboratory, is based on an equivalent level of categorization.

ABX and AX Discrimination

Many speech perception studies do not evaluate discrimination. When discrimination is evaluated, a roving-level ABX procedure is the most common technique employed. In the ABX technique, the subject is presented on each trial with two stimuli (A and B), then asked if a third (X) is equivalent to A or B. This task has a great deal of face validity in that both of the stimuli being compared (A and B) are presented during the trial. Macmillan et al. (1977) have provided a theoretical analysis of ABX, with a basis for comparison with more standard psychophysical tasks. While there is some evidence that experienced subjects may be able to ignore the A stimulus and perform the ABX task as a modified BX (or same-different task), naive subjects seem to respond on the basis of X being more like A or B (in a standard SD or AX task, $A = X$ or $A < X$, whereas in an ABX task, $B = X$ or $B < X$ or $B > X$). One excellent example of the difference in results from these types of tasks can be found in the research on auditory temporal acuity. For diotic stimuli, practiced subjects in an AX task can report differences in stimulus onset for differences of 2 msec or less, but require approximately 18 msec to indicate which stimulus had the earlier onset in an ABX task. Naive subjects tend to exhibit only the 18 msec difference for both tasks. Hirsh and Sherrick (1961) have argued that the smaller detection threshold may be

based on spectral correlates of the temporal difference, rather than on subjects responding directly to the stimulus onset difference (see Pastore, 1987b, for a more detailed summary of this literature).

Other Psychophysical Procedures

A roving-level oddity procedure sometimes has been employed to measure discrimination. On each trial the subject is presented with three or four stimulus tokens, with one of the tokens different from the other identical tokens. The task of the subject is to indicate which of the stimuli was different. The oddity procedure also seems to have face validity, but it does not lend itself easily to a theoretical analysis, and thus the results from this procedure are difficult to compare with those from more standard procedures (Macmillan et al., 1977).

The two-interval forced-choice (2IFC) procedure is a common one in the auditory detection and discrimination literature. The procedure is relatively insensitive to the criterion differences and has a theoretical basis for comparison with other psychophysical procedures (Green and Swets, 1974; Egan, 1975). In the recognition version of this procedure, a subject is presented with two different stimuli on each trial and is asked which of the stimuli is greater along the specified dimension. A roving-level 2IFC procedure has been used successfully to investigate temporal order discrimination for complex nonspeech stimuli, producing discrimination results similar to roving-level AX and ABX procedures under equivalent conditions (Pastore et al., 1987).

More Standard Psychophysical Procedures

Adaptive psychophysical procedures have been successfully employed in the study of speech categories and of complex sounds believed to be possibly related to speech perception. Summerfield (1981) used PEST (parameter estimation by sequential testing, Taylor and Creelman, 1967) with speech stimuli to determine consonant boundaries. Pastore et al. (1982) used the Levitt (1971) up-down procedure with a 2IFC task and 2:1 rule to determine temporal-order thresholds for simple and complex analog to speech stimuli.

Selective Adaptation

Selective adaptation techniques initially were employed in the speech perception literature to demonstrate the existence of specialized feature detectors (discussed below in the section on models). The basic notion was that different speech categories are each mediated by a specialized feature detector. Stimuli drawn from specific continua each activate different feature detectors, and stimuli from the boundaries between feature detectors may sometimes activate one or the other feature detector. If one of the feature detectors is adapted or fatigued, it will become less responsive and result in a lower probability of activation by boundary stimuli (and thus a higher probability of activation of the alternative category). Selective adaptation typically involves use of a labeling procedure (and/or, with less frequency, a discrimination procedure) to measure the location of the category boundary with no adaptation and following repeated exposure to (adaptation by) a specific stimulus. The effectiveness of a stimulus as an adaptor was believed to indicate the degree to which a stimulus is an example of a given category. Alternative explanations of selective adaptation that do not require the assumption of feature detectors include stimulus contrast effects (Diehl, Kluender, and Parker, 1985; Diehl, 1981; Sawusch and Jusczyk, 1981;

Sawusch and Millennix, 1985), range effects (Parducci, 1974), criterion shifts (Warren, 1985; Warren and Meyers, 1987), and altered or mutated organizational units that contribute to the perceptual whole (Warren and Meyers, 1987).

Selective adaptation procedures have been used to evaluate the relationship between speech cues (Ades, 1974), as well as between tones varying in temporal onset and speech stimuli varying in voice onset time (Pisoni, 1980). Miller et al. (1983) used selective adaptation to measure the relative strength of category membership for stimuli drawn from a given speech continuum.

Reaction Time Measures

Reaction time measures are quite common in the cognitive sciences literature and often have been used in the speech perception literature. In general, disjunctive reaction time measures tend to be longer for stimuli near a labeling boundary than for stimuli drawn from within labeling categories (e.g., Pisoni and Tash, 1974). Differences in reaction time results have been used as indicators of integral versus separable cues for category membership (e.g., Wood, 1976; Pastore et al., 1976) and to evaluate differences in feature integration (Massaro, 1987a).

Scaling

The multidimensional scaling (MDS) technique is a powerful analysis tool that, if carefully and knowledgeably employed, can provide a representation of the physical stimuli in terms of a type of perceptual space and an estimation of the number and basic nature of relevant perceptual dimensions. Unfortunately, very few studies have been based on multidimensional scaling of the similarities or differences among speech stimuli. Shepard (1972) provided a MDS analysis of the Miller and Nicely (1955) consonant confusion data, while Soli, Arabie, and Carroll (1986) used INDCLUS (*in*dividual *d*ifferences *clus*tering) to provide a new MDS analysis of these same data.

Comparison Across Procedures

In the definition of categorical perception for speech stimuli (see below), discrimination performance for stimuli drawn from within a given category must be at or near chance. Discrimination typically is measured with an ABX procedure, although sometimes an oddity procedure is used which yields equivalent results. When an AX procedure is used with relatively naive subjects, discrimination performance maintains the general pattern found with ABX, although performance tends to be higher (Pisoni, 1977). When subjects are practiced with speech and nonspeech stimuli varying in relative onset time (VOT or TOT) under minimal uncertainty AX conditions (discrimination only between two stimuli), discrimination performance exhibits a Weber's law type of relationship, with highest performance at minimum onset differences (Kewley-Port et al., 1987). This change in the discrimination performance may well be due to the learned ability of subjects to use subtle stimulus cues that would tend to be ignored as inconsistent or unreliable cues under high uncertainty conditions (for further discussion, see the section on practice effects, and also Hirsh and Sherrick, 1961; Pastore et al., 1982). A theoretically interesting condition exists when discrimination performance is equivalent under minimal-uncertainty and high-uncertainty conditions, since the basis of high-uncertainty discrimination performance must be relatively simple and can be subjected to careful analysis.

CATEGORICAL PERCEPTION

Categorical perception (CP) has a very precise definition in the auditory perception literature. Its demonstration requires both labeling and discrimination tasks for stimuli drawn from a given physical continuum. CP is said to occur under the following conditions: (1) a sharp, stable labeling boundary between two perceived categories, (2) a peak in discrimination at the labeling boundary, (3) troughs of chance performance within labeling categories (allowing for local discrimination peaks), and (4) a high correlation between the empirical discrimination performance and discrimination performance predicted from the labeling results (Studdert-Kennedy et al., 1970). Demonstration of only the sharp labeling boundary was designated a category boundary effect (Wood, 1976; Pastore, 1981). In the early 1970s, CP was believed to be unique to speech and to represent the absolute recoding of the continuously variable speech signal into discrete perceptual (phonetic) categories.

In the middle and late 1970s, a number of findings significantly altered this conceptualization of CP for auditory stimuli. CP was found for several nonspeech acoustic continuations: (a) sawtooth rise time (Cutting and Rosner, 1974, 1976—although see Rosen and Howell, 1981, and Cutting, 1982); (b) noise onset time (Miller et al., 1976); (c) masked tones (Pastore et al., 1977); and (d) musical intervals (Burns and Ward, 1978; Pastore et al., 1983). This demonstrates that CP is not unique to speech stimuli. The second change in our understanding of CP concerned the assertion that it represented absolute recoding into discrete perceptual categories. Initial research had demonstrated chance discrimination performance (typically with a roving-level ABX procedure) for stimuli drawn from the same category and separated by one or two (arbitrarily defined) steps along the given physical continuum being manipulated. However, each category typically spanned more than two stimulus steps, and discrimination for three-step differences often was better than chance, while one- and two-step discrimination often was better than predicted from labeling results, assuming absolute recoding of stimulus information. This problem with the notion of absolute categorization was ignored until discrimination performance was demonstrated to be better than chance when subjects were given practice with the given stimuli (Samuel, 1977; Carney et al., 1977). Apparently because CP was no longer considered unique to speech, speech perception researchers have tended in recent years to focus on other phenomena that seem to be more unique to speech. The continuing importance of CP is the demonstration that under high uncertainty conditions some stimulus continua exhibit a high correlation between discrimination and categorization, especially for stimuli located near category boundaries. If CP is a general property of perceiving complex auditory stimuli, then it will be critical that researchers develop an understanding of the basis for CP in terms of the roles played by perceptual thresholds and perceptual learning.

Categorical perception is more broadly, and less precisely, defined in the subfields of animal and infant auditory psychophysics, as in the field of cognitive sciences. In these fields the term *categorical perception* seems to be used as a euphemism for categorization of stimuli. Harnad (1987) has chapters by recognized researchers from across the broad spectrum of cognitive sciences and thus represents an excellent summary of categorization research across sensory, perceptual, and cognitive modalities.

MODELS OF CATEGORIZATION AND CATEGORICAL PERCEPTION

Modern models of categorical perception fall into two general conceptual categories: exemplar and boundary models. In the late 1960s and early 1970s, a third class of models, neural feature detector models, was popular (e.g., Eimas and Corbit, 1973), but the notion that the categorical nature of speech perception is due to highly specialized, automatically

responding feature detectors seems to have fallen into disfavor as empirical studies demonstrated flexibility in speech categories (for a discussion, see Remez, 1987a; Diehl, 1987). While exemplar and boundary models are presented as being mutually exclusive, it is quite possible that both types of processes are important in categorization, and that different forms of each of these types of processes may have different degrees of importance.

Exemplar or Token Models

Exemplar or token models conjecture that there is a set of ideal stimuli for each category. The exemplars need not ever be realized as actual stimuli. Actual stimuli are compared with the exemplar stimuli, and categorization is based on some measure of similarity between the actual and ideal stimuli. Although specification of the nature of the comparison process is lacking in the auditory literature, it has been developed in the cognitive sciences literature and is briefly discussed in the next section.

The work of Stevens and Blumstein (1978, 1981) on cues for the perception of place of articulation is an excellent example of the exemplar approach, wherein certain spectral characteristics of stimuli at onset are believed to distinguish the different perceived categories correlated with changes in place of articulation.

The motor theory of speech perception is a different type of exemplar model in which the exemplar is defined in terms of the characteristics of the source of the stimulus, and not directly by the spectral characteristics of the acoustic stimuli. According to motor theory, we possess some form of internalized knowledge of exemplars of the articulations for each discrete speech category, and that perception is based on an evaluation of the type of articulation that might have produced the given sounds (Studdert-Kennedy et al., 1970; Liberman and Mattingly, 1985; Repp and Liberman, 1987). This knowledge may be based on some idealized token or prototype (which never can be achieved), or some form of perceptual norm that is unique to speech and represents an internalization of the production conventions of the listener's language. Most researchers on speech perception have not dealt with the specific nature of the prototype or with the process by which a perceiver reaches the decision of category membership for the given sound; Chistovitch (1985) is a notable exception. Finally, questions concerning critical dimensions, weighting of dimension importance, and perceptual distance metrics are important issues that have not been very thoroughly investigated.

The model of vowel perception proposed by James D. Miller (1987) attempts to map the relevant stimulus dimensions for various vowels and proposes a perceptual decision mechanism based on dynamic properties of the stimulus. This model could be considered to be more a boundary or threshold model (discussed next) than an exemplar model.

The fuzzy logical model of Massaro (1987a, 1987b) is an exemplar-type model for speech perception that deals with the decision process. In many respects this model is similar to many cognitive models of categorical behavior (and is discussed below along with the cognitive models).

Boundary or Threshold Models

Boundary models are based on the notion that there are qualitative changes in perceptual quality along stimulus continua, with categorization for certain types of stimuli based on the specific combinations of perceptual quality. An example of a boundary conceptualization of categorization can be found in the research on temporal order identification or judgment (TOJ). Hirsh (1959), Miller et al. (1976), Pisoni (1977), and Pastore et al. (1982)

all conjectured that limitations on the ability of listeners to identify the onset order of acoustic stimuli may serve as a basis for the categorical perception of speech stimuli varying along a voice onset time (VOT) continuum. Listeners require such an onset difference of approximately 18 msec to identify which of two stimuli had an earlier onset, although they can reliably detect an onset difference at 2 msec or less. According to this specific boundary hypothesis, the perception of voicelessness in speech requires (at least in part) that the subjects be able to perceive an unvoiced component of the stimulus prior to the onset of voicing. Criticism of this specific hypothesis as a valid explanation of voicing categorization can be found in the work of Summerfield (1982) and Rosen and Howell (1987b). Pastore (1987b) provides a more detailed evaluation of categorical perception in terms of thresholds, while Macmillan (1987) provides a detection-theory analysis in terms of the Durlach-Braida model (see above).

Categorization Research for Speech

Early investigations of speech perception focused on the relationship between the physical properties of the speech signal and perception of speech categories. By mapping the physical stimulus dimensions and cues correlated with a given speech category, this basic research provided fundamental knowledge of categorization consistent with both exemplar and boundary models of categorization.

Although the most common models for speech perception are exemplar, most recent research on speech perception has tended to focus on the location of the category boundary. For instance, cross-language studies tend to focus on differences in boundary location across languages, or on the relative influence of specific speech cues on a given type of boundary location. Trading relations, the phenomena now sometimes claimed to be unique to speech (Repp, 1982; Repp and Liberman, 1987; Pisoni and Luce, 1987), are demonstrations of two cues operating together or in opposition in altering a given boundary location. This focus on category boundaries probably is based on the relative ease with which boundaries can be measured, and research using boundary location as the dependent measure certainly does indicate changes in category membership.

There are several notable exceptions to this focus on boundaries. The work of Stevens and Blumstein (1981), and research motivated by their work, has attempted to identify invariant characteristics of the speech signal that may cue the perception of specific speech categories and thus, in essence, to provide specification of the critical exemplar properties for speech. A second exception is the attempt of Miller et al. (1983) to measure the strength of category membership of within-category stimuli based on the relative magnitude of selective adaptation effect on the category boundary location.

Multidimensional scaling techniques are powerful analysis tools that could provide an indication of the nature of exemplars in terms of the clustering of perceived stimuli within categories and the relative perceptual distances among stimuli. However, this statistical tool has been little used in the auditory categorization literature.

Ecological (Gibsonian) Theory

The ecological theory of perception, originally developed by the late J.J. Gibson, represents a relatively new and different approach to the study of perception (Gibson, 1966, 1976). For the Gibsonian, the organism is part of the environment with which it interacts. Perception is direct, is the consequence of the organism interacting with its environment, and is the means by which the organism maintains contact with its environment. The organism

perceives objects in its environment and the relevant qualities of the object relevant for the organism (the affordances). The organism does not perceive the stimuli and does not somehow compute a representation of the objects in its environment from the stimulus properties. Sound stimuli "ordinarily provide information about what produced them and where the source is located" (Jenkins, 1985). This does not mean that stimulus properties should be ignored, but rather that the stimulus properties should be directly related to the properties of the objects and that those object properties are altered in a meaningful manner. The ecological researcher should study "what aspects of the environment are perceivable by ear, and (secondly) what acoustic dimensions are the carriers of (or the information for) these audible properties of the environment" (VanDerveer, 1979). A Gibsonian might well criticize most modern studies of speech perception and categorical perception for having employed stimulus continua that are not direct functions of articulation continua, and thus lacking in ecological validity.

Most ecologically oriented researchers work with the visual modality or with movement, although there have been a few auditory studies. Jenkins (1985) provides an excellent summary of the relevance and value of the ecological approach to understanding the nature of acoustic information. Warren has provided an ecological analysis of auditory perception for breaking and bouncing events (Warren and Verbrugge, 1984; Warren, Kim, and Husney, 1987). Rosenblum has provided an ecological analysis of the perception of moving acoustic events (Rosenblum, Carello, and Pastore, 1987). The study of the perception of hand clapping by Repp (1987) seems to be motivated by the Gibsonian emphasis on the use of ecologically valid stimuli and the identification of source characteristics. Although the findings in the Repp study were largely negative, this research does represent one of the few solid studies that attempts to apply to new, natural situations the techniques and procedures developed to study speech stimuli.

Fowler has been a strong advocate of considering ecological validity in the study of speech, and her research certainly reflects this strong theoretical orientation (Fowler, 1980, 1983, 1984). The advantage of Fowler's approach to understanding speech is that the central, and seemingly insolvable, problems of identifying the invariant acoustic properties that, from the perspective of more traditional researchers, are the basis of the categorization and segmentation of speech perception, are simply not relevant (Diehl, 1986, provides a critical review of this approach).

There are strong theoretical reasons why categorization of acoustic stimuli and categorical perception of acoustic events apparently have not been addressed by ecologically oriented research. If subjects are directly perceiving the quality of the objects and events in their environment based on the sounds they produce, then the relevant issue is whether the perceived categories accurately reflect categories of the physical events or objects, and not whether they reflect categories of the acoustic stimuli or the dimensions of those acoustic stimuli. Acoustic stimuli should be studied in terms of the dynamic flow of changing information correlated with changes in the object and its location within the context of acoustic information about absence of change in the environment.

Cognitive Science Modeling

In cognitive sciences, both stimulus specification and theoretical detail tend to be less precise than in the auditory perception literature. However, cognitive scientists have attempted to deal with general issues concerning the structure of natural and learned categories.

The structures of natural categories tend to be complex and poorly defined. Those natural categories that are considered to be well defined typically have a critical set of features that are individually necessary, but only jointly sufficient to define category membership (Katz and Postal, 1964). Rather, natural categories usually are defined in terms of typical, rather than critical features. Once the typical features have been identified, how does the observer use this information to determine category membership?

There appear to be two main classes of feature-based models (Estes, 1986). Prototype models are equivalent to the exemplar models described above. According to prototype modes the observer stores some form of an abstract exemplar or representation of each category. Category membership then is based on some form of evaluation of the perceived similarity between a given stimulus and the prototype or exemplar. The use of multidimensional scaling techniques to identify central tendencies for category membership seems obvious for such models. Feature validity models assume that information about category features is stored and then used in evaluating category membership. Information about a given feature may include mean feature, range or dispersion of feature values, relative frequency of occurrence, relative importance, etc. Feature validity models differ from each other in terms of what type of feature information is stored and how feature information is actually used in evaluating category membership. Independent cue feature models assume that each feature is evaluated separately, with the comparison results combined in an additive fashion to judge category membership. Interactive cue feature models assume some form of relative, conditional, or conjoined evaluation of feature properties (see Medin, Dewey, and Murphy, 1983, for a description and discussion of these various types of models). Massaro's fuzzy logical model of perception (FLMP) is a type of feature integration model that results in the categorization of perceived stimuli (Oden and Massaro, 1978; Massaro, 1987a, 1987b). According to this model, there are three stages of analysis. During the first stage information is transduced by the sensory systems and various features are derived in an independent and continuous fashion. The second stage combines feature information and then evaluates these features against "perceptual-unit definitions, or prototypes" in terms of complex, arbitrary fuzzy logical propositions. This fuzzy logical evaluation reflects the degree to which the comparison is valid (not the probability of occurrence), and the importance of each feature is greater when other features are low in importance. In this model, the prototypes seem to be special types of interactive cue feature specifications of perceptual classes or categories. In the final pattern classification stage, the summed merits of each potential prototype are evaluated relative to all others in a manner similar to Luce's (1959) choice rule. Massaro has used his FLMP to study the categroization of speech stimuli.

References and Bibliography

Ades, A.
1974 How phonetic is selective adaptation? Experiments on syllable position and vowel environment. *Perception and Psychophysics* 16:61–67.

Ahumada, A., and J. Lovell
1971 Stimulus features in signal detection. *Journal of the Acoustical Society of America* 49:1751–1756.

Ahumada, A., R. Marken, and A. Sandusky
1975 Time and frequency analyses of auditory signal detection. *Journal of the Acoustical Society of America* 57:385–390.

Aiken, E.G.
1969 Auditory discrimination learning: Prototype storage and distinctive feature detection mechanism. *Perception and Psychophysics* 6:95–96.

Aldrich, W.M., and S.J. Barry
1980 Critical bandwidth measured by the pulsation threshold technique. *Journal of Auditory Research* 20:137–141.

Anderberg, M.R.
1973 *Cluster Analysis for Applications.* New York: Academic Press.

Ashby, F.G., and J.T. Townsend
1986 Varieties of perceptual independence. *Psychology Review* 93:154–179.

Auerbach, C., and P.A. Sperling
1974 A common visual-auditory space: Evidence for its reality. *Perception and Psychophysics* 16:129–135.

Bacon, S.P., and J.F. Brandt
1982 Auditory processing of vowels by normal-hearing and hearing-impaired listeners. *Journal of Speech and Hearing Research* 25:339–347.

Ballas, J.A., and J.H. Howard
1978a Predicting Individual Listener Confusions in the Classification of Complex, Steady-State Sounds. ONR Contract Number N00014-75-C-0308. ONR-78-8. December.
1978b Two Approaches to Category Representation in Aural Classification. ONR Contract Number N00014-75-C 0308. ONR-78-9. December.

Bashford, J.A., and R.M. Warren
1987 Multiple phonemic restorations follow the rules for auditory induction. *Perception and Psychophysics* 42:114–121.

Bennett, K.B., R. Parasuraman, J.H. Howard, and A.J. O'Toole
1984 Auditory induction of discrete tones in signal detection tasks. *Perception and Psychophysics* 35:570–578.

Berg, B.G., and D.E. Robinson
1987 Multiple observations and internal noise. *Journal of the Acoustical Society of America* 81:S33 (abstract).

Berger, K.W.
1964 Some factors in the recognition of timbre. *Journal of the Acoustical Society of America* 35:1888–1891.

Berkely, D.

1987 Localisation in rooms. In W.A. Yost and G. Gourevitch (Eds.), *Directional Hearing*. New York: Springer-Verlag.

Berliner, J.E., and N.I. Durlach

1973a Intensity perception IV: Resolution in roving-level discrimination. *Journal of the Acoustical Society of America* 53:1270–1287.

1973b A perceptual anchor model for context coding in intensity perception. *Journal of the Acoustical Society of America* 54:336A.

Berliner, J.E., N.I. Durlach, and L.D. Braida

1977 Intensity perception VII: Further data on roving-level discrimination and the resolution of bias edge effects. *Journal of the Acoustical Society of America* 61:1577–1585.

1978 Intensity perception IX: Effect of fixed standard on resolution in identification. *Journal of the Acoustical Society of America* 64:687–689.

Bilsen, F.A., and J.L. Goldstein

1974 Pitch of dichotically delayed noise and its possible spectral basis. *Journal of the Acoustical Society of America* 55:292–296.

Blackith, R.E., and Reyment, R.A.

1971 *Multivariate Morphometrics*. New York: Academic.

Bladon, A.

1982 Arguments against formants in the auditory representation in speech. Pp. 95–102 in R. Carlson and B. Granstrom (Eds.), *The Representation of Speech in the Peripheral Auditory System*. Amsterdam: Elsevier Biomedical Press.

Bladon, R.A., and B. Lindblom

1981 Modeling the judgment of vowel quality differences. *Journal of the Acoustical Society of America* 65(5):1414–1422.

Blauert, J.

1982 *Spatial Hearing*. Cambridge, Mass.: MIT Press.

Bock, H.H.

1974 *Automatiches Klassifikation*. Gottingen, W. Germany: Vandenhoeck Und Ruprecht.

Bond, Z.S.

1976 On the specification of input units in speech perception. *Brain and Language* 3:72–87.

Bookshire, R.H.

1972 Visual and auditory sequencing by aphasic subjects. *Journal of Communication Disorders* 5:259–269.

Braida, L.D., and N.I. Durlach

1972 Intensity perception II: Resolution in one-interval paradigms. *Journal of the Acoustical Society of America* 51:483–502.

1986 Peripheral and central factors in intensity perception. In G.M. Edelman, W.E. Gall, and W.M. Cohen (Eds.), *Functions of the Auditory System*. New York: Wiley.

Braida, L.D., J.S. Lim, J.E. Berliner, N.I. Durlach, W.M. Rabinowitz, and S.R. Purks

1984 Intensity perception XII: Perceptual anchor model of context coding. *Journal of the Acoustical Society of America* 76:722–731.

Bregman, A.S.

1978a Auditory streaming: Competition among alternative organisations. *Perception and Psychophysics* 23:391–398.

1978b The formation of auditory streams. In J. Requin (Ed.), *Attention and Performance VII*. Hillsdale, N.J.: Lawrence Erlbaum.

1978c Asking the "What for" question in auditory perception. In M. Kubovy and J. Pomerantz (Eds.), *Perceptual Organization*. Hillsdale, N.J.: Lawrence Erlbaum Press.

1980 The conceptual basis of perception and action. In *Perception and Cognition II: Presentations on Art Education Research*. Montreal: Concordia University.

1982 Two-f. ~tor theory of auditory organization. *Journal of the Acoustical Society of America* 72:S10 A.

1987 The meaning of duplex perception: Sounds as transparent objects. Pp. 95–111 in M.E.H. Schouten (Ed.), *The Psychophysics of Speech Perception*. Boston, Mass.: Martinus Nijhoff.

Bregman, A.S., and J. Campbell

1971 Primary auditory stream segregation and the perception of order in rapid sequences of tones. *Journal of Experimental Psychology* 89:244–249.

Bregman, A.S., and G.L. Dannenbring

1973 The effect of continuity on auditory stream segregation. *Perception and Psychophysics* 13:308–312.

1977 Auditory continuity and amplitude edges. *Canadian Journal of Psychology* 31:151–159.

Bregman, A.S., and S. Pinker
1978 Auditory streaming and the building of timbre. *Canadian Journal of Psychology* 32:19–31.

Bregman, A.S., and A.I. Rudnicky
1975 Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception and Performance* 1:263–267.

Bregman, A.S., J. Abramson, and C. Darwin
1985 Spectral integration based upon common amplitude modulation. *Perception and Psychophysics* 37:483–493.

Broadbent, D.E., and P. Ladefoged
1959 Auditory perception of temporal order. *Journal of the Acoustical Society of America* 31:1539–1540.

Brubaker, B.S., and R.M. Warren
1987 Detection of infratonal repetition of frozen noise: Singularity recognition or pattern recognition? *Journal of the Acoustical Society of America* 82:5–93.

Bukofzer, M.F.
1947 *Music in the Baroque Era*. New York: Norton.

Burns, E.M., and W.D. Ward
1978 Categorical perception—Phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America* 63:456–468.

Bush, R.R., R.D. Luce, and R.M. Rose
1964 Learning models in psychophysics. Pp. 201–217 in R.C. Atkinson (Ed.), *Mathematical Psychology*. Stanford, Calif.: Stanford University Press.

Butler, R.A.
1975 The influence of the external and middle ear on auditory discrimination. In D.L. Neff and W.D. Keidel (Eds.), *Handbook of Sensory Physiology*. New York: Springer-Verlag.
1985 An analysis of monaural displacement of sound in space. *Perception and Psychophysics* 41:1–7.

Carmon, A., and I. Nachshon
1971 Effect of unilateral brain damage on perception of temporal order. *Cortex* 7:410–418.

Carney, A.E., G.P. Widin, and N.F. Viemeister
1977 Non-categorical perception of stop consonants. *Journal of the Acoustical Society of America* 62:961–970.

Carroll, J.D., and J.B. Kruskal
1978 Multidimensional scaling. Pp. 892–907 in J. Tanur and W. Kruskal (Eds.), *International Encyclopedia of Statistics*. New York: MacMmilan and the Free Press.

Casseday, J.H., and E. Covey
1987 Central auditory pathways in directional hearing. In W.A. Yost and G. Gourevitch (Eds.), *Directional Hearing*. New York: Springer-Verlag.

Chaney, R.B., and J.C. Webster
1966 Information in certain multidimensional sounds. *Journal of the Acoustical Society of America* 40:447–455.

Charbonneau, G.R.
1981 Timbre and the perceptual effects of three types of data reduction. *Computer Music Journal* 5(2):10–19.

Chase, S., P. Bugnacki, L.D. Braida, and N.I. Durlach
1983 Intensity perception XII: Effect of presentation probability on absolute identification. *Journal of the Acoustical Society of America* 73:279–284.

Cheng, P.W., and R.G. Pachella
1984 A psychophysical approach to dimensional separability. *Cognitive Psychology* 16:279–304.

Cherry, E.C.
1953 Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America* 25:975–979.

Cherry, E.C., and R. Wiley
1967 Speech communications in very noisy environments. *Nature* 214:1164.

Chistovitch, L.A.
1985 Central auditory processing of peripheral vowel spectra. *Journal of the Acoustical Society of America* 77:789–805.

Chowning, J.M.
1980 *Computer Synthesis of the Singing Voice, in Sound Generation in Winds, Strings, Computers*. Stockholm: Royal Swedish Academy of Music.

Clavita, F.B., F.V. Sseligo, and S.D. Zimmer
1974 Temporal pattern discrimination in cats with insular-temporal lesions. *Brain Research* 79:153–156.
Cohen, E.A.
1984 Some effects of inharmonic partials on interval perception. *Music Perception Journal* 1:323–349.
Cohen, M.F., and E.D. Schubert
1987 The effect of cross-spectrum correlation on the detectability of a noise band. *Journal of the Acoustical Society of America* 81:721–723.
Colburn, H.S., and N.I. Durlach
1978 Models of binaural interaction. Pp. 467–520 in E.C. Carterette and M.P. Friedman (Eds.), *Handbook of Perception: Hearing*. New York: Academic Press.
Cole, R.A., and B. Scott
1973 Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology* 27:441–449.
Colquhoun, W.P.
1975 Evaluation of auditory, visual, and dual-mode displays for prolonged sonar monitoring in repeated sessions. *Human Factors* 17:425–437.
Coninx, F.
1978 The detection of combined differences in frequency and intensity. *Acustica* 39:137.
Corcoran, D.W.J., A. Carpenter, J.C. Webster, and M.M. Woodhead
1968 Comparisons of training techniques for complex sound identification. *Journal of the Acoustical Society of America* 44:157–167.
Cowan, N.
1984 On short and long auditory stores. *Psychological Bulletin* 96:341–370.
Coxon, A.P.M., and P.M. Davies
1982 *User's Guide to Multidimensional Scaling: With Reference to MDS (X) Library of Computer Programs.* Exeter, N.Y.: Heinemann.
Cramer, E.M., and W.H. Huggins
1958 Creation of pitch through binaural interaction. *Journal of the Acoustical Society of America* 30:413–417.
Creelman, C.D.
1960 Detection of signals of uncertain frequency. *Journal of the Acoustical Society of America* 32:805–810.
Cullinan, W.L., E. Erdos, R. Schaefer, and M.E. Tekiell
1977 Perception of temporal order of vowels and consonant-vowel syllables. *Journal of Speech and Hearing Research* 20:742–751.
Cutting, J.E.
1982 Plucks and bows are categorically perceived, sometimes. *Perception and Psychophysics* 31:462–476.
Cutting, J.E. and B.S. Rosner
1974 Catagories and boundaries in speech and music. *Perception and Psychophysics*, 16:564–570.
1976 Discrimination functions predicted from categories in speech and music. *Perception and Psychophysics*, 20:87–88.
Dannenbring, G.L.
1976 Perceived auditory continuity with alternately rising and falling frequency transitions. *Canadian Journal of Psychology* 30:99–114.
Dannenbring, G.L., and A.S. Bregman
1976 Stream segregation and the illusion of overlap. *Journal of Experimental Psychology: Human Perception and Performance* 2:544–555.
de Boer, B.
1976 "Residue" and auditory pitch perception. In *Handbook of Sensory Physiology*. Berlin: Springer-Verlag.
Demany, L.
1985 Perceptual learning in frequency discrimination. *Journal of the Acoustical Society of America* 78:1118–1120.
Deutsch, D.
1974 An auditory illusion. *Nature* (London) 251:307–309.
1982 *The Psychology of Music*. New York: Academic Press.
Dewson, J.H.
1964 Speech sounds discrimination by cats. *Science* 144:555–556.

Dewson, J.H., and A. Cowey
1969 Discrimination of auditory sequences by monkeys. *Nature* 222:695–697.
Diehl, R.L.
1981 Feature detectors for speech: A critical reappraisal. *Psychological Bulletin* 89:1–18.
1986 Coproduction and direct perception of phonetic segments: A critique. *Journal of Phonetics* 14:61–66.
1987 Auditory constraints on speech perception. Pp. 210–219 in M.E.H. Schouten (Ed.), *The Psychophysics of Speech Perception*. Boston: Martinus Nijhoff.
Diehl, R.L., and K.R. Kluender
1987 On the categorisation of speech sounds. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition*. New York: Cambridge University Press.
Diehl, R.L., K.R. Kluender, and E.M. Parker
1985 Are selective adaptation and contrast effects really distinct?. *Journal of Experimental Psychology: Human Perception and Performance* 11:209–220.
Divenyi, P.L., and I.J. Hirsh
1972 Pitch changes in trills and vibrato. *Journal of the Acoustical Society of America* 51:A138.
1974 Identification of temporal order in three-tone sequences. *Journal of the Acoustical Society of America* 56:144–151.
1978 Some figural properties of auditory patterns. *Journal of the Acoustical Society of America* 64:1369–1385.
Dorman, M.F., J.E. Cutting, and L.J. Raphael
1975 Perception of temporal order in vowel sequences with and without formant transitions. *Journal of Experimental Psychology: Human Perception and Performance* 104:121–129.
Dowling, W.J.
1968 Rhythmic fission and perceptual organisation. *Journal of the Acoustical Society of America* 44:369.
1973 The perception of interleaved melodies. *Cognitive Psychology* 5:322–337.
Duifhuis, H., L.F. Willems, and R. J. Sluyter
1982 Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception. *Journal of the Acoustical Society of America* 71:1568–1580.
Durlach, N.I.
1972 Binaural signal detection: Equalisation and cancellation theory. In J. Tobias (Ed.), *Foundations of Modern Auditory Theory*. New York: Academic Press.
Durlach, N.I., and L.D. Braida
1969 Intensity perception I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America* 46:372–383.
Durlach, N.I., and H.S. Colburn
1978 Binaural phenomena. In E.C. Carterette and Friedman (Eds.), *Handbook of Perception (Hearing)*. New York: Academic Press.
Efron, R.
1963 Temporal perception. *Aphasia and Brain* 86:403–424.
1973 Conservation of temporal information by perceptual systems. *Perception and Psychophysics* 14:518–530.
Egan, J.P.
1975 *Signal Detection Theory and ROC Analysis*. New York: Academic Press.
Egan, J.P., G.Z. Greenberg, and A.I. Schulman
1961 Interval of the time uncertainty in auditory detection. *Journal of the Acoustical Society of America* 33:771–778.
Egan, J.P., A.I. Schulman, and G.Z. Greenberg
1961 Memory for waveform and time uncertainty in auditory detection. *Journal of the Acoustical Society of America* 33:770–781.
Egeth, H., and R. Pachella
1969 Multidimensional stimulus identification. *Perception and Psychophysics* 5:341–346.
Ehret, G.
1987 *Categorical Perception of Sound Signals: Facts and Hypotheses from Animal Studies*. Harnad Press.
Eimas, P.D., and J.D. Corbit
1973 Selective adaptation of linguistic feature detectors. *Cognitive Psychology* 4:99–109.
Elfner, L.F.
1969 Continuity in alternately sounded tone and noise signals in a free field. *Journal of the Acoustical Society of America* 46:914–917.

1971 Continuity in alternately sounded tonal signals in a free field. *Journal of the Acoustical Society of America* 49:447–449.

Elfner, L.F., and W.E. Caskey

1965 Continuity effects with alternately sounded noise and tone signals as a function of manner of presentation. *Journal of the Acoustical Society of America* 38:543–547.

Elfner, L.F., and J.L. Homick

1966 Some factors affecting the perception of continuity in alternately sounded tone and noise signals. *Journal of the Acoustical Society of America* 40:27–31.

1967a Continuity effects with alternately sounding tones under dichotic presentation. *Perception and Psychophysics* 2:34–36.

1967b Auditory continuity effects as a function of the duration and temporal location of the interpolated signal. *Journal of the Acoustical Society of America* 42:576–579.

Eriksen, C.W., and H.W. Hake

1955 Multidimensional stimulus differences and accuracy of discrimination. *Journal of Experimental Psychology* 50:153–160.

Estes, W.K.

1950 Toward a statistical theory of learning. *Psychology Review* 57:94–107.

1986 Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology* 115:155–174.

Fastl, H.

1975 Pulsation patterns of sinusoids vs. critical band noise. *Perception and Psychophysics* 18:95–97.

1978 Frequency discrimination for pulsed vs. modulated tones. *Journal of the Acoustical Society of America* 63:275–277.

Fay, R.R., and A. Feng

1987 Localization in non-mammalian species. In W.A. Yost and G. Gourevitch (Eds.), *Directional Hearing*. New York: Springer–Verlag.

Fay, W.H.

1966 *Temporal Sequence in the Perception of Speech.* The Hague: Mouton.

Feth, L.L.

1972 Combinations of amplitude and frequency differences in auditory discrimination. *Acustica* 26:67.

Fitts, P.M., J.R. Peterson, and G. Wolpe

1963 Cognitive aspects of information processing: II. Adjustments to stimulus redundancy. *Journal of Experimental Psychology* 65:507–514.

Flege, J.E., and J. Hillenbrand

1987 Limits on phonetic accuracy in foreign language speech production. In Ioup and Weinberger (Eds.), *Interlanguage Phonology: The Acquisition of a Second Language Sound System.* Cambridge, Mass.: Newbury House.

Foulke, E., and T.G. Sticht

1969 Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin* 72:50–62.

Fowler, C.A.

1980 Coarticulation and theories of extrinsic timing. *Journal of Phonetics* 8:113–133.

1981a Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research* 24:127–139.

1981b A relationship between coarticulation and compensatory shortening. *Phonetica* 38:35–50.

1983 Realism and unrealism: A reply. *Journal of Phonetics* 11:303–322.

1984 Segmentation and coarticulated speech in perception. *Perception and Psychophysics* 36:359–368.

1986 An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14:3.28.

Fraisse, P.

1963 *The Psychology of Time* (J. Leith, Trans.). New York: Harper & Row.

Freed, D.J.

1987 Perceptual Control of Timbre in Musical Applications Using Psychophysical Functions. Masters Thesis, Northwestern University.

Friedman, M.P., E.C. Carterette, L. Nakatani, and A. Ahumada

1968 Comparisons of some learning models for response bias in signal detection. *Perception and Psychophysics* 3:5–11.

Gardner, R.B., and C.J. Darwin
1986 Grouping of vowel harmonics by frequency modulation: Absence of effects on phonemic categorisation. *Perception and Psychophysics* 40(3):183–187.

Garner, W.R.
1951 The accuracy of counting repeated short tones. *Journal of Experimental Psychology* 41:310–316.
1962 Uncertainty and structure as psychological concepts. New York: Wiley Press.
1970 The stimulus in information processing. *American Psychology* 25:350–358.

Garner, W.R., and C.D. Creelman
1964 Effect of redundancy and duration on absolute judgments of visual stimuli. *Journal of Experimental Psychology* 67:168–172.

Garner, W.R., and G.L. Fefoldy
1970 Integrability of stimulus dimensions in various types of information processing. *Cognitive Psychology* 1:225–241.

Garner, W.R., and R.L. Gottwald
1967 Some perceptual factors in the learning of sequential patterns of binary events. *Journal of Verbal Learning and Verbal Behavior* 6:582–589.
1968 The perception and learning of temporal patterns. *Quarterly Journal of Experimental Psychology* 20:97–109.

Garner, W.R., and H.W. Hake
1951 The amount of information in absolute judgments. *Psychology Review* 58:446–459.

Garner, W.R., and J. Morton
1969 Perceptual independence: Definitions, models, and experimental paradigms. *Psychological Bulletin* 72:233–259.

Garrett, K.L., and E.C. Healy
1987 An acoustic analysis of fluctuations in the voices of normal adult speakers across three times a day. *Journal of the Acoustical Society of America* 82:58–62.

Gati, I., and A. Tversky
1982 Representations of qualitative and quantitative dimensions. *Journal of Experimental Psychology: Human Perception and Performance* 8:325–340.

Getty, D.J., J.A. Swets, J.B. Swets, and D.M. Green
1979 On the prediction of confusion matrices from similarity judgments. *Perception and Psychophysics* 26:1–19.

Gibson, E.J.
1969 *Principles of Perceptual Learning and Development.* New York: Century–Crofts.

Gibson, J.J.
1966 *The Senses Considered as Perceptual Systems.* Boston: Houghton Mifflin Press.
1974 A note on ecological optics. In E.C. Carterette and M.P. Friedman (Eds.), *Handbook of Perception: Historical and Philosophical Roots of Perception.* New York: Academic Press.
1976 *The Ecological Approach to Visual Perception.* Boston: Houghton Mifflin Press.

Gibson, J.J., and E.J. Gibson
1955 Perceptual learning: Differentiation or enrichment. *Psychology Review* 62:32–41.

Gilkey, R.H.
1987 Spectral and temporal comparisons in auditory masking. Pp. 26–36 in W.A. Yost and C.S. Watson (Eds.), *Auditory Processing of Complex Sounds.* Hillsdale, N.J.: Lawrence Erlbaum Press.

Gilkey, R.H., and D.E. Robinson
1986 Models of auditory masking: A molecular psychophysical approach. *Journal of the Acoustical Society of America* 79:1499–1510.

Gilkey, R.H., D.E. Robinson, and T.E. Hanna
1985 Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise. *Journal of the Acoustical Society of America* 78:1207–1219.

Gilliam, J.D., and W.M. Mills
1976 Information extraction from contralateral cues in the detection of signals of uncertain frequency. *Journal of the Acoustical Society of America* 59:1428.

Glasberg, B.R., B.C.J. Moore, and R.A. Lutfi
1982 Off-frequency listening and masker uncertainty. *Journal of the Acoustical Society of America* 72:273–275.

Glasberg, B.R., B.C.J. Moore, and I. Nimmo-Smith
1984 Comparison of auditory filter shapes derived with three different maskers. *Journal of the Acoustical Society of America* 75:536–544.

Goldberg, R., N.A. Macmillan, and L.D. Braida
1985 A perceptual-anchor interpretation of categorical phenomena on a vowel continuum. *Journal of the Acoustical Society of America* 77(SI):S7(A).

Goldstein, J.L.
1967 Auditory spectral filtering and monaural phase perception. *Journal of the Acoustical Society of America* 41:193.

Golledge, R.G., and J.N. Rayner
1982 *Proximity and Preference: Problems in the Multidimensional Analysis of Large Data Sets.* Minneapolis: University of Minnesota Press.

Graham, N., and J. Nachmias
1971 Detection of grating patterns containing two spatial frequencies: A comparison of single-channel and multiple channel models. *Vision Research* 11:251–259.

Gravetter, F., and G.R. Lockhead
1973 Critical range as a frame of reference for stimulus judgment. *Psychology Review* 80:203–216.

Green, D.M.
1961 Detection of auditory sinusoids of uncertain frequency. *Journal of the Acoustical Society of America* 33:897–903.
1964 Consistency of auditory detection judgments. *Psychological Review* 71:392–407.
1971 Temporal auditory acuitv. *Psychological Review* 78(b):542–551.
1988 *Profile Analysis.* London: Oxford Press.

Green, D.M., and T.G. Birdsall
1978 Detection and recognition. *Psychological Review* 85:192–206.

Green, D.M., and G.B. Henning
1969 Audition. In *Annual Review of Psychology.* Palo Alto, Calif.: Annual Reviews Inc.

Green, D.M., and R.D. Luce
1974 Counting and timing mechanisms in auditory discrimination and reaction time. In Krantz (Ed.), *Contemporary Developments in Mathematical Psychology.* San Francisco, Calif.: Freeman Press.
1976 Variability of magnitude estimates: A timing theory analysis. *Perception and Psychophysics* 15:291–300.

Green, D.M., and J.A. Swets
1974 *Signal Detection Theory and Psychophysics.* San Francisco: Peninsular Press.

Green, D.M., and D.L. Weber
1980 Detection of temporally uncertain signals. *Journal of the Acoustical Society of America* 67:1304.

Green, D.M., and W.A. Yost
1975 Binaural analysis. In D.L. Neff and W.D. Keidel (Eds.), *Handbook of Sensory Physiology: Hearing.* Netherlands: Springer-Verlag.

Green, D.M., G. Kidd, and M.C. Picardi
1983 Successive versus simultaneous comparison in auditory intensity discrimination. *Journal of the Acoustical Society of America* 73:639–643.

Green, D.M., R.D. Luce, and A.F. Smith
1980 Individual magnitude estimates for various distributions of signal intensity. *Perception and Psychophysics* 27:483–488.

Green, D.M., C.R. Mason, and G. Kidd
1984 Profile analysis: Critical bands and duration. *Journal of the Acoustical Society of America* 74:1163–1167.

Green, D.M., D.L. Weber, and J.E. Duncan
1977 Detection and recognition of pure tones in noises. *Journal of the Acoustical Society of America* 62:948–954.

Green, P.E., and F.J. Carmone
1970 *Multidimensional Scaling and Related Techniques in Marketing Analysis.* Boston: Allyn & Bacon.

Green, P.E., and V.R. Rao
1972 *Applied Multidimensional Scaling: A Comparison of Approaches and Algorithms.* New York: Holt Press.

Greenewalt, C.H.
1968 *Bird Song: Acoustics and Physiology.* Washington, D.C.: Smithsonian Institution Press.

Greenspan, S.L., H.C. Nusbaum, and D.B. Pisoni
1986 Perceptual learning of synthetic speech produced by rule. Pp. 43–85 in *Research on Speech Perception Progress.* Report No. 12. Indiana University.

Grey, J.M.

1977 Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America* 61(5):1270–1277.

Grey, J.M., and J.W. Gordon

1978 Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America* 63(5):1493–1500.

Grey, J.M., and J.A. Moorer

1977 Perceptual evaluations of synthesised musical instrument tones. *Journal of the Acoustical Society of America* 62(2):454–462.

Grunke, M.E., and D.B. Pisoni

1982 Some experiments on perceptual learning of mirror image acoustic patterns. *Perception and Psychophysics 31:210–218.*

Gundy, R.F.

1961 Auditory detection of an unspecified signal. *Journal of the Acoustical Society of America* 33: 1008–1012.

Guttman, N., and B. Julesz

1963 Lower limit of auditory periodicity analysis. *Journal of the Acoustical Society of America* 35:610.

Hanna, T.E.

1984 Discrimination of reproducible noise as a function of bandwidth and duration. *Perception and Psychophysics* 36:409–416.

Hanna, T.E. and D.E. Robinson

1985 Phase effects for a sine wave masked by reproducible noise. *Journal of the Acoustical Society of America* 77:1129–1140.

Harnad, S.

1987 *Categorical Perception: The Groundwork of Cognition.* New York: Cambridge University Press.

Harris, J.D.

1952 The decline of pitch discrimination with time. *Journal of Experimental Psychology* 43:93–99.

Hall, J.W., III

1987 Experiments on comodulation masking release. In W.A. Yost and C.S. Watson (Eds.), *Auditory Processing of Complex Sounds.* Hillsdale, N.J.: Lawrence Erlbaum.

Hartman, E.B.

1954 The influence of practice and pitch-distance between tones on the absolute identification of pitch. *American Journal of Psychology* 67:1–14.

Hartmann, W.M.

1983 Localisation in rooms. *Journal of the Acoustical Society of America* 74:1380–1391.

1985 Perceptual entities from complex inharmonic tones (Abstract). P. 168 in *Proceedings of the Association for Research in Otolaryngology.* 8th Meeting.

1987 Pitch perception and the organisation and integration of auditory entities. In G.W. Edelman, W.E. Gall, and W.M. Cowan (Eds.), *The Auditory System Development, Physiology, Psychophysics.* New York: John Wiley.

Hartmann, W.M., and G.M. Hnath

1982 Detection of mixed modulation. *Acustica* 50:297–312.

Hartmann, W.M., and M.A. Klein

1980 Theory of frequency modulation detection for low modulation frequencies. *Journal of the Acoustical Society of America* 67:935.

1981 The effect of uncertainty on the detection of frequency modulation at low modulation rates. *Perception and Psychophysics* 30:417–424.

Hartmann, W.M., and K.A. Long

1976 Time dependence of pitch perception—vibrato experiment. *Journal of the Acoustical Society of America* 60:5-42.

Heise, G.A., and G.A. Miller

1951 An experimental study of auditory patterns. *American Journal of Psychology* 64:68–77.

Helmholtz, H.

1855 *On the Sensations of Tone* (Trans. A.J. Ellis). New York: Dover Press.

Henning, G.B.

1966 Frequency discrimination of random-amplitude tones. *Journal of the Acoustical Society of America* 39:336–339.

Hillenbrand, J.

1987 Speaker-independent vowel classification based on fundamental frequency and format frequencies. *Journal of the Acoustical Society of America* 81:S93.

Hirsh, I.J.

1959 Auditory perception of temporal order. *Journal of the Acoustical Society of America* 31:759–767.

1976 Order of events in three modalities. In S.K. Hirsh, D.H. Eldridge, I.J. Hirsh, and S.R. Silverman (Eds.), *Hearing and Davis: Essays Honoring Hallowell Davis*. St. Louis: Washington University Press.

Hirsh, I.J., and C.R. Sherrick

1961 Perceived order in different sense modalities. *Journal of Experimental Psychology* 62:423–432.

House, A.S., K.N. Stevens, T.T. Sandel, and J.B. Arnold

1962 On the learning of speech-like vocabularies. *Journal of Verbal Learning and Verbal Behavior* 1:133–143.

Houtgast, T.

1972 Psychophysical evidence for lateral inhibition in hearing. *Journal of the Acoustical Society of America* 51:1885–1894.

1973 Psychophysical experiments on "tuning curves" and "two-tone" inhibition. *Acustica* 29:168–179.

1974a Masking patterns and lateral inhibition. Pp. 258 265 in E. Zwicker and E. Terhardt (Eds.), *Facts and Models in Hearing*. Berlin: Springer-Verlag.

1974b The slopes of masking patterns. Pp. 269–272 in E. Zwicker and E. Terhardt (Eds.), *Facts and Models in Hearing*. Berlin: Springer-Verlag.

Houtsma, A.J.M.

1979 Musical pitch of two-tone complexes and predictions by modern pitch theories. *Journal of the Acoustical Society of America* 66:87–99.

1981 Pitch of unequal amplitude dichotic two-tone complexes. *Journal of the Acoustical Society of America* 69:1778–1785.

Houtsma, A.J.M., N.I. Durlach, and L.D. Braida

1980 Intensity perception XI: Experimental results on the relation of intensity resolution to loudness matching. *Journal of the Acoustical Society of America* 68:807–813.

Houtsma, A.J.M., N.I. Durlach, and D.M. Horowitz

1987 Comparative learning of pitch and loudness identification. *Journal of the Acoustical Society of America* 81:129–132.

Houtsma, A.J.M., and J.L. Goldstein

1972 The central origin of the pitch of complex tones: Evidence from musical interval recognition. *Journal of the Acoustical Society of America* 51:520–529.

Howard, J.H.

1976 The Psychophysical Structure of Complex Underwater Sounds. ONR Contract Number N00014-75-C-0308. ONR-76-2.

1977 Psychophysical structure of eight complex underwater sounds. *Journal of the Acoustical Society of America* 62:149–156.

1982 Feature selection in human auditory perception. *Pattern Recognition* 15:397–403.

Howard, J.H., and J.A. Ballas

1978a Classification of Amplitude-Modulated Noise Patterns with Extended Practice. ONR Contract Number N00014-75-C-0308. ONR-78-7. December.

1978b Feature Selection in Auditory Perception. ONR Contract Number N00014-75-C-0308. ONR-78-5. July.

1980 Syntactic and semantic factors in the classification of non-speech transient patterns. *Perception and Psychophysics* 28:431–439.

1981 Feature selection in auditory perception. In D.J. Getty and J.H. Howard (Eds.), *Auditory and Visual Pattern Recognition*. Hillsdale, N.J.: Lawrence Erlbaum.

1982 Acquisition of acoustic pattern categories by exemplar observation. *Organismic Behavior and Human Performance* 30:157–173.

1983 Perception of simulated propeller cavitation. *Human Factors* 25:643–656.

Howard, J.H., and J.J. O'Hare

1984 Human classification of complex sounds. *Naval Research Reviews* 1:26–32.

Howard, J.H., and E.B. Silverman

1976 A multidimensional scaling analysis of 16 complex underwater sounds. *Perception and Psychophysics* 19:193–200.

Howard, J.H., J.A. Ballas, and D.C. Burgy

1978 Feature Extraction and Decision Processes in the Classification of Amplitude-Modulated Noise Patterns. ONR Contract Number N00014-75-C-0308. ONR-78-4. July.

Howard, J.H., A.J. O'Toole, R. Parasuraman, and K.B. Bennett

1984 Pattern-detection in uncertain frequency detection. *Perception and Psychophysics* 35:256–264.

Howard, J.H., A.J. O'Toole, and S.E. Rice
1986 The role of frequency versus informational cues in uncertain frequency detection. *Journal of the Acoustical Society of America* 79:788–791.

Hulse, S.H.
1987 Comparative aspects of complex acoustic perception. In W.A. Yost and C.S. Watson (Eds.), *Auditory Processing of Complex Sounds*. Hillsdale, N.J.: Lawrence Erlbaum.

Jamieson, D.G.
1987 Studies of possible psychoacoustic factors underlying speech perception. Pp. 220–231 in M.E.H Schouten (Ed.), *The Psychophysics of Speech Perception*. Boston: Nijhoff-Martinus.

Jardine, N., and R. Sibson
1971 *Mathematical Taxonomy*. New York: Wiley Press.

Jassem, W.
1979 Classification of fricative spectra using statistical discriminant functions. Pp. 77–91 in B. Lindblom and S. Ohman (Eds.), *Frontiers of Speech Communication Research*. London: Academic Press.

Jeffress, L.A.
1972 Binaural signal detection: Vector theory. In J. Tobias (Ed.), *Foundations of Modern Auditory Theory*. New York: Academic Press.

Jenkins, J.J.
1985 Acoustic information for objects, places, and events. In W.H. Warren and R.E. Shaw (Eds.), *Persistence and Change: Proceedings of the First International Conference on Event Perception*. Hilldale, N.J.: Lawrence Erlbaum.

Jesteadt, W., and R.C. Bilger
1974 Intensity and frequency discrimination in one- and two-interval paradigms. *Journal of the Acoustical Society of America* 55:1266–1274.

Jesteadt, W., R.D. Luce, and D.M. Green
1977 Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance* 3:92–104.

Johnson, D.M., and E.R. Hafter
1980 Uncertain frequency detection: Cuing and condition observation. *Perception and Psychophysics* 28:143–149.

Joos, M.
1948 Acoustic phonetics. Supplement to *Language* (Language Monograph No. 23) 24:1–136.

Julesz, B. and I.J. Hirsh
1972 Visual and auditory perception: An essay of comparison. Pp. 283–340 in E.D. David and P.B. Denes (Eds.), *Human Communication: A Unified View*. New York: McGraw-Hill.

Jusczyk, P.W., D.B. Pisoni, M.A. Reed, A. Fernald, and M. Myers
1983 Infants's discrimination of the duration of a rapid spectrum change in nonspeech signals. *Science* 222:175–177.

Jusczyk, P.W., D.B. Pisoni, A. Walley, and J. Murray
1980 Discrimination of relative onset time of two component tones by infants. *Journal of the Acoustical Society of America* 67:262–270.

Karnichaya, E.G., V.N. Mushnikov, and S. Zhukov
1975 Auditory processing of steady-state vowels. Pp. 37–53 in G. Fant and M.A.A. Tatham (Eds.), *Auditory Analysis and Perception of Speech*. London: Academic Press.

Katz, J.J., and P.M. Postal
1964 *An Integrated Theory of Linguistic Descriptions*. Cambridge, Mass.: MIT Press.

Kay, R.H., and D.R. Matthews
1972 On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated yones. *Journal of Physiology* 225:657–677.

Kendall, G.S., and W.L. Martens
1984 Simulating the cues for spatial hearing in natural environments. *Proceedings of the 1984 International Music Conference*. Paris.

Kerkhof, G.A., T.W. van der Schaaf, and H.J. Korving
1980 Auditory signal detection: Effect of long term practice and time on task. *Perception and Psychophysics* 28:79–81.

Kewley-Port, D. and D.B. Pisoni
1984 Identification and discrimination of rise time: Is it categorical or non-categorical. *Journal of the Acoustical Society of America* 75:1168–1176.

Kewley-Port, D., C.S. Watson, and D.C. Foyle
1987 Auditory Temporal Acuity in Relation to Category Boundaries: Speech and Non-Speech Stimuli. Manuscript in Preparation, Indiana University.
Kidd, G., C.R. Mason, and D.M. Green
1986 Auditory profile analysis of irregular sound spectra and non-speech stimuli. *Journal of the Acoustical Society of America* 79:1045–1053.
Kinchla, R., and F. Smyzer
1967 A diffusion model of perceptual memory. *Perception and Psychophysics* 2:219–229.
Kinney, J.A.S.
1961 Discrimination of auditory and visual patterns. *American Journal of Psychology* 74:529–541.
Klein, M.A.
1980 Perception of Frequency Modulation Width at Low Modulation Frequencies. Thesis, Michigan State University.
Klein, M.A., and W.M. Hartmann
1981 Binaural edge pitch. *Journal of the Acoustical Society of America* 70:51–61.
Klein, W., R. Plomp, and L.C.W. Pols
1970 Vowel spectra, vowel spaces, and identification. *Journal of the Acoustical Society of America* 48(4/2):999–1009.
Knudsen, E.I., M. Konishi, and J.D. Pettigrew
1977 Receptive fields of auditory neurons in the owl. *Science* 198:1278–1280.
Kobus, D., A. Russotti, J. Schlitchting, C. Haskell, G. Carpenter, and J. Wojtowicz
1986 Multimodal detection and recognition performance of sonar operators. *Human Factors* 28:23–30.
Koffka, K.
1935 *Principles in Gestalt Psychology.* New York: Harcourt and Brace.
Kohler, W.
1929 *Gestalt Psychology.* New York: Horace Liveright.
Konishi, M.
1986 Centrally synthesized maps of sensory space. *Trends in Neuroscience* 9:163–168.
Kornbrot, D.E.
1980 Attention bands: Some implications for categorical judgment. *British Journal of Math and Statistics in Psychology* 33:1–16.
Kronberg, H., V. Mellert, and C. Schreiner
1974 Dichotic and monaural pulsation thresholds. *Proceedings of the 8th International Congress on Acoustics* (London) 1:143.
Kruskal, J.B.
1964a Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1–27.
1964b Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29:115–129.
Kruskal, J.B., and M. Wish
1978 *Multidimensional Scaling.* Beverly Hills, Calif.: Sage Publications.
Kubovy, M.
1987 Concurrent pitch segregation. In W.A. Yost and C.S. Watson (Eds.), *Auditory Processing of Complex Sounds.* Hillsdale, N.J.: Lawrence Erlbaum.
Kubovy, M., J.E. Cutting, and R.M. McGuire
1974 Hearing with the third ear: Dichotic perception of a melody without monaural familiarity cues. *Science* 186:272–274.
Kubovy, M., and R. Jordan
1979 Tone segregation by phase: On the phase sensitivity of the single ear. *Journal of the Acoustical Society of America* 66:100–106.
Kuhl, P., and J.D. Miller
1974 Discrimination of speech sounds by the chinchilla: /t/ vs. /d/ in CV syllables. *Journal of the Acoustical Society of America* 56:S52 (abstract).
1978 Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America* 63:905–917.
Kuhn, G.
1987 Physical variables for localization. In W.A. Yost and G. Gourevitch (Eds.), *Directional Hearing.* New York: Springer–Verlag.
Kuwada, S., and T.C. Yin
1987 Physiological mechanisms of localization. In W.A. Yost and G. Gourevitch (Eds.), *Directional Hearing.* New York: Springer-Verlag.

Ladefoged, P., and A. Traill
1984 Linguistic phonetic descriptions of clicks. *Language* 60:1–20.
Larkin, W., and G.Z. Greenberg
1970 Selective attention in uncertain frequency detection. *Percpetion and Psychophysics* 8:179–184.
Law, H.G.
1984 *Research Methods for Multimode Data Analysis*. New York: Praeger.
Layton, B.
1975 Differential effects of two non-speech sounds on phonemic restoration. *Bulletin of the Psychonomic Society* 6:487–490.
Leek, M.R., and C.S. Watson
1984 Learning to detect auditory pattern components. *Journal of the Acoustical Society of America* 76(4):1037–1044.
Leek, M.R., M.F. Dorman, and Q. Summerfield
1987 Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America* 81(1):148–154.
Levitt, H.
1971 Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America* 49:467–477.
Libby, E.R.
1980 *Binaural Hearing an Amplification*. Chicago: Zerotron.
Liberman, A.M., and I.G. Mattingly
1985 The motor theory of speech perception revised. *Cognition* 21:1–36.
Lichte, W.H.
1941 Attributes of complex tones. *Journal of Experimental Psychology* 28(6):455–480.
Lim, J.S., W.M. Rabinowits, L.D. Braida, and N.I. Durlach
1977 Intensity perception VIII: Loudness comparisons between different types of stimuli. *Journal of the Acoustical Society of America* 62:1577–1587.
Lindner, W.A.
1968 Recognition performance as a function of detection criterion in a simultaneous detection-recognition task. *Journal of the Acoustical Society of America* 44:204–211.
Lippmann, R.P., L.D. Braida, and N.I. Durlach
1976 Intensity perception V: Effect of payoff matrix on absolute identification. *Journal of the Acoustical Society of America* 59:129–134.
Lockhead, G.R.
1966 Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology* 72:95–104.
1970 Identification and the form of multidimensional discrimination space. *Journal of Experimental Psychology* 85:1–10.
1972 Processing dimensional stimuli: A note. *Psychology Review* 79:410–419.
Long, G.R.
1973 The role of the standard in an auditory amplitude discrimination experiment. *Perception and Psychophysics* 13:49–59.
Luce, R.D.
1959 *Individual Choice Behavior*. New York: John Wiley.
Luce, R.D., D.M. Green, and D.L. Weber
1976 Attention bands in absolute identification. *Perception and Psychophysics* 20:49–54.
Luce, R.D., R.M. Nosofsky, D.M. Green, and A. Smith
1982 The bow and sequential effects in absolute identification. *Perception and Psychophysics* 32(5):397–408.
Lundeen, C., and A.M. Small
1980 Detection of signals with identical energy spectra but different waveforms under conditions of signal uncertainty. *Journal of the Acoustical Society of America* 67(7):210 (L).
Mackie, R.R., C.D. Wylie, R.R. Ridihalgh, T.E. Shults, and M. L. Seltser
1981 *Some Dimensions of Auditory Sonar Signal Perception and Their Relationships to Target Classification* (Technical Report 2723-1). (AD A102598). Goleta, Calif.: Human Factors Research, Inc.
Macmillan, N.A.
1983 A psychophysical interpretation of a "categorical perception" experiment by Hary and Massaro. *Perception and Psychophysics* 34:494–498.
1985 What sort of psychophysics is infant psychophysics? In Trehube and B. Schneider (Eds.), *Auditory Development in Infancy*. New York: Plenum.

1987 Beyond the categorical/continuous distinction: A psychophysical approach to processing modes. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition.* New York: Cambridge University Press.

Macmillan, N.A., L. Braida, and R.F. Goldberg

1987 Central and peripheral processing in the perception of speech and nonspeech sounds. Pp. 46–65 in M.E.H. Schouten (Ed.), *The Psychophysics of Speech Perception.* Boston: Martinus-Nijhoff.

Macmillan, N.A., H. Kaplan, and C.D. Creelman

1977 The psychophysics of categorical perception. *Psychology Review* 84:452–471.

Maiwald, D.

1967a Die Berechnung der Modulationsschwellen mit Hilfe eines Funktionsschemas. *Acustica* 18:193.

1967b Beziehungen Zwischen Schallspektrum, Mithorschwelle und der Erregung des Gehors. *Acustica* 18:69.

1967c Ein Funktionsschema des Gehors zur Beschreibung der Erkennbarkeit Kleiner Frequenzund Amplitudenderunger. *Acustica* 18:81.

Mann, V.A., and B.H. Repp

1980 Influence of vocalic context on perception of the /t/–/s/ distinction. *Perception and Psychophysics* 28:213–228.

Marley, A.A., and V.T. Cook

1984 A fixed rehearsal capacity interpretation of limits on absolute identification performance. *British Journal of Math and Statistics in Psychology* 37:136–151.

1986 A limited capacity rehearsal model for psychophysical judgments applied to magnitude estimation. *Journal of Math and Psychology* 30:339–390.

Martens, J.P.

1981 Audibility of harmonics in a periodic complex. *Journal of the Acoustical Society of America* 70:234–237.

1984 Comment on 'algorithm for extraction of pitch and pitch salience from complex tonal signals.' *Journal of the Acoustical Society of America* 75:626–628.

Massaro, D.W.

1987a Categorical partition: A fuzzy-logical model of categorization. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition.* New York: Cambridge University Press.

1987b Psychophysics versus specialized processes in speech perception: An alternative perspective. Pp. 46–65 in M.E.H. Schouten (Ed.), *The Psychophysics of Speech Perception.* Boston: Martinus-Nijhoff.

Mathews, M.V., and J.R. Pierce

1980 Harmony and non-harmonic partials. *Journal of the Acoustical Society of America* 68:1252–1257.

Mazur, J.E., and R. Hastie

1978 Learning as accumulation: A reexamination of the learning curve. *Psychological Bulletin* 85:1256–1274.

McAdams, S.

1982 Spectral fusion and the creation of auditory images. Pp. 279–289 in M. Clynes (Ed.), *Music, Mind, and Brain: The Neuropsychology of Music.* New York: Plenum Press.

1984a Spectral Fusion, Spectral Parsing, and the Formation of Auditory Images. Ph.D Thesis, Stanford University.

1984b The auditory image: A metaphor for musical and psychological research on auditory organization. Pp. 289–324 in W.R. Crozier and A.J. Chapman (Eds.), *Cognitive Processes in the Perception of Art.* North Holland: Amsterdam Press.

McAdams, S., and K. Saariaho

1985 Qualities and functions of musical timbre. Pp. 367–374 in B. Traux (Ed.), *Proceedings of the International Computer Music Conference.* San Francisco: Computer Music Association.

McFadden, D.

1975 Masking and the binaural system. In D. Tower (Ed.), *The Nervous System, Human Communication and Its Disorders.* New York: Raven Press.

1987 Comodulation detection differences using noise-band signals. *Journal of the Acoustical Society of America* 81(5):1519–1527.

Medin, D.L., and L.W. Barsalou

1987 Categorization processes and categorical perception. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition.* New York: Cambridge University Press.

Medin, D.L., G.I. Dewey, and T.... Murphy

1983 Relationships between category learning: Evidence that abstraction is not automatic. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9:607–625.

Meyer, M.L.
1899 Is memory of absolute pitch capable of development by training? *Psychology Review* 6:514–516.
Middlebrooks, J.C., and E.I. Knudsen
1984 A neural code for spatial tuning in the cat's superior colliculus. *Journal of Neuroscience* 4:2621–2634.
Miller, G.A.
1956 The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychology Review* 63:81–97.
Miller, G.A., and W.R. Garner
1944 Effect of random presentation on the psychometric function: Implications for a quantal theory of discrimination. *American Journal of Psychology* 57:451467.
Miller, G.A., and G.A. Heise
1950 The trill threshold. *Journal of the Acoustical Society of America* 22:637–638.
Miller, G.A., and J.C.R. Licklider
1950 The intelligibility of interrupted speech. *Journal of the Acoustical Society of America* 2:167–173.
Miller, G.A., and P.E. Nicely
1955 An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27:338–352.
Miller, J.D.
1987 Auditory-perceptual processing of speech waveforms. In W.A. Yost and C.S. Watson (Eds.), *Auditory Processing of Complex Sounds*. Hillsdale, N.J.: Lawerence Erlbaum.
Miller, J.D., C.C. Wier, R.E. Pastore, W.J. Kelly, and R. Dooling
1976 Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America* 60:686–696.
Miller, J.L.
1981 Some effects of speaking rate on phonetic perception. *Phonetica* 38:159–180.
1986 Limits on later occurring rate information for phonetic perception. *Language and Speech* 29:13–24.
Miller, J.L., C.M. Connine, T.M. Schermer, and K.R. Kluender
1983 A possible auditory basis for internal structure of phonetic categories. *Journal of the Acoustical Society of America* 73:2124–2133.
Miller, J.L. and P.D. Eimas
1982 Feature detectors and speech perception: A critical evaluation. Pp. 111–145 in D.G. Albrecht (Ed.), *Recognition of Pattern and Form*. New York: Springer Verlag.
Miller, J.L., and A.M. Liberman
1979 Some effects of later occurring information on the perception of stop consonant and semivowel. *Perception and Psychophysics* 25:457–465.
Miller, J.R., and E.C. Carterette
1975 Perceptual space for musical structures. *Journal of the Acoustical Society of America* 58(3):711–720.
Mills, A.W.
1972 Auditory localization. In J. Tobias (Ed.), *Modern Foundations in Auditory Theory*. New York: Academic Press.
Monahan, J.S., and G.R. Lockhead
1977 Identification of integral stimuli. *Journal of Experimental Psychology: General*. 1:94–110.
Montague, W.E.
1965 Effect of irrelevant information on a complex auditory discrimination task. *Journal of Experimental Psychology* 69:230–236.
Moore, B.C.J.
1982 *An Introduction to the Psychology of Hearing*. Second Edition. London: Academic Press.
Moore, B.C.J., B.R. Glasberg, and R.W. Peters
1985b Relative dominance of individual partials in determining the pitch of complex tones. *Journal of the Acoustical Society of America* 77:1853–1860.
Moore, B.C.J., R.W. Peters, and B.R. Glasberg
1985a Thresholds for the detection of inharmonicity in complex tones. *Journal of the Acoustical Society of America* 77:1861–1867.
Moorer, J.A., J.M. Grey, and J. Strawn
1978 Lexicon of analyzed tones—Part three: The trumpet. *Computer Music Journal* 2(2):23–31.
Morgan, B.J., M.M. Woodhead, and J.C. Webster
1976 On the recovery of physical dimensions of stimuli, using multidimensional scaling. *Journal of the Acoustical Society of America* 60:186–189.

Musicant, A.D., and R.A. Butler
1984 The influence of pinnae-based spectral cues on sound localisation. *Journal of the Acoustical Society of America* 75:1196–1200.
Neff, D.L., and B.P. Callaghan
1987 Simultaneous masking by small numbers of sinusoids under conditions of uncertainty. In W.A. Yost and C.S. Watson (Eds.), *Auditory Processing of Complex Sounds*. Hillsdale, N.J.: Lawrence Erlbaum.
Neff, D.L., and D.M. Green
1987 Masking produced by spectral uncertainty with multicomponent maskers. *Perception and Psychophysics* 41:409–415.
Neisser, U.
1967 *Cognitive Psychology*. New York: Appelton.
1976 *Cognition and Reality*. San Francisco: Freeman.
Neisser, U., and W. Hirst
1974 Effect of practice on the identification of auditory sequences. *Perception and Psychophysics* 15:391–398.
Nickerson, R.S., and B. Freeman
1974 Discrimination of the order of the components of repeating tone sequences: Effects of frequency separation and extensive practice. *Perception and Psychophysics* 16:471–477.
Norman, D.A.
1967 Temporal confusions and limited capacity processors. *Acta Psychologica* 27:293–297.
Nosofsky, R.M.
1983a Shifts of attention in the identification and discrimination of intensity. *Perception and Psychophysics* 33:103–112.
1983b Information integration and the identification of stimulus noise and critical noise in absolute judgment. *Journal of Experimental Psychology: Human Perception and Performance* 9:299–309.
1985 Overall similarity and the identification of separable-dimension stimuli. *Perception and Psychophysics* 38:415–432.
1986 Attention, similarity, and the identification categorization relationship. *Perception and Psychophysics* 115:39–57.
Oden, G.C., and D.W. Massaro
1978 Integration of featural formation in speech perception. *Psychology Review* 85:686–696.
O'Hanlon, J.F., A.E. Schmidt, and C.H. Baker
1965 Doppler discrimination and the effect of a visual alertness indicator upon detection of auditory sonar signals in a sonar watch. *Human Factors* 7:129–140.
Ohman, S.
1967 Numerical model of coarticulation. *Journal of the Acoustical Society of America* 41:310–320.
O'Shaughnessy, D.
1986 Speaker recognition. *IEEE ASSP* (10–86):4–17.
Ortmann, 0.
1926 On the melodic relativity of tones. *Psychological Monographs* 35(1):Whole No. 162.
Parducci, A.
1974 Contextual effects: A range-frequency analysis. In E.C. Carterette and M.P. Friedman (Eds.), *Handbook of Perception*. New York: Academic Press.
Parker, E.M., and R.L. Diehl
1984 Identifying vowels in CVC syllables: Effects of inserting and noise. *Perception and Psychophysics* 36:369–380.
Parker, E.M., R.L. Diehl, and K.R. Kluender
1986 Trading relations in speech and nonspeech. *Perception and Psychophysics* 39:129–142.
Pastore, R.E.
1981 Possible psychoacoustic factors in speech perception. In P.D. Emias and J.L. Miller (Eds.), *Perspectives on the Study of Speech*. Hillsdale, N.J.: Lawrence Erlbaum.
1983 Temporal order judgment of auditory stimulus offset. *Perception and Psychophysics* 33:54–62.
1987a Categorical perception: Some psychophysical models. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition*. New York: Cambridge University Press.
1987b Possible acoustic bases for the perception of voicing contrasts. Pp. 188–198 in M.E.H. Schouten (Ed.), *The Psychophysics of Speech Perception*. Boston: Martinus-Nijhoff.
Pastore, R.E., L.B. Harris, and J.K. Kaplan
1982 Temporal order identification: Some parameter dependencies. *Journal of the Acoustical Society of America* 71:430–436.

Pastore, R.E., W.A. Ahroon, K.J. Buffuto, C.D. Friedman, J.S. Puleo, and E.A. Fink
1977 Common-factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance* 3:686–696.

Pastore, R.E., W.A. Ahroon, J.S. Puleo, D.B. Crimmins, L. Golowner, and R.S. Berger
1976 Processing interaction between two dimensions of nonphonetic auditory signals. *Journal of Experimental Psychology: Human Perception and Performance* 2:267–276.

Pastore, R.E., J. Kaplan-Layer, C.B. Morris, and R.J. Logan
1987 Temporal order identification for tone/noise stimuli with onset transitions. *Perception and Psychophysics* (under review).

Pastore, R.E., R. Logan, C. Morris, and J.K. Layer
1986 Perceptual Learning Effects in the Perception of Speech-Like Stimuli. American Psychological Association Meeting, August 1986, Washington D.C.

Pastore, R.E., M.A. Schmuckler, L. Rosenblum, and R. Szczesiul
1983 Duplex perception with musical stimuli. *Perception and Psychophysics* 33:469–474.

Patterson, J.H., and D.M. Green
1970 Discrimination of transient signals having identical energy spectra. *Journal of the Acoustical Society of America* 48:894–905.

Patterson, R.D.
1973 The effects of relative phase and the number of components on residue pitch. *Journal of the Acoustical Society of America* 53:1565–1572.

Patterson, R.D., and G.B. Henning
1977 Stimulus variability and auditory filter shape. *Journal of the Acoustical Society of America* 62:649.

Penner, M.J.
1972 The effect of pay-offs and cue tones on detection of sinusoids of uncertain frequency. *Perception and Psychophysics* 11:198–202.

Pfafflin, M.S., and M.V. Mathews
1966 Detection of auditory signals in reproducible noise. *Journal of the Acoustical Society of America* 39:340–345.

Pfafflin, S.M.
1968 Detection of auditory signals in reproducible noise. *Journal of the Acoustical Society of America* 43:487–490.

Pisoni, D.B.
1971 On the Nature of Categorical Perception of Speech Sounds. Ph.D. Dissertation, University of Michigan.
1973 Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics* 13:253–260.
1975 Auditory short-term memory and vowel perception. *Memory and Cognition* 3:7–18.
1977 Identification and discrimination of the relative onset times of two-component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America* 61:1352–1361.
1980 Adaptation of the relative onset time of two component tones: Implications for voicing perception in stops. *Perception and Psychophysics* 28:337–346.
1987 Auditory perception of complex sounds: Some comparisons of speech vs. nonspeech signals. In W.A. Yost and C.S. Watson (Eds.), *Auditory Processing of Complex Sounds*. Hillsdale, N.J.: Lawrence Erlbaum.

Pisoni, D.B., and J.H. Lazarus
1974 Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America* 55:328–333.

Pisoni, D.B., and P.A. Luce
1987 Trading relations, acoustic cue integration, and context effects in speech perception. Pp. 155–172 in M.E.H. Schouten (Ed.), *The Psychophysics of Speech Perception*. Boston: Martinus-Nijhoff.

Pisoni, D.B., and J. Tash
1974 Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics* 15:285–290.

Pisoni, D.B., T.D. Carrell, and S.J. Gans
1983 Perception of duration of rapid spectrum changes in speech and nonspeech signals. *Perception and Psychophysics* 34:314–322.

Pisoni, D.B., R.N. Aslin, A.J. Perey, and B.L. Hennessy
1982 Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance* 8:293–314.
Piston, W.
1947 *Counterpoint.* New York: Norton Press.
Plomp, R.
1964 The ear as a frequency analyzer. *Journal of the Acoustical Society of America* 36:1628–1636.
1967 Timbre as a multidimensional attribute to complex interaural relationships. *Journal of the Acoustical Society of America* 59:1197.
1976 *Aspects of Tone Sensation: A Psychophysical Study.* London: Academic Press.
Plomp, R., and A.M. Mimpin
1968 The ear as a frequency analyzer II. *Journal of the Acoustical Society of America* 43:764–767.
Pollack, I.
1952 The information of elementary auditory displays. *Journal of the Acoustical Society of America* 24:745–749.
1953 The information of elementary auditory displays II. *Journal of the Acoustical Society of America* 25:765–769.
1972 Memory for auditory waveform. *Journal of the Acoustical Society of America* 52:1209–1215.
1975a Identification of random auditory waveforms. *Journal of the Acoustical Society of America* 58:1262–1271.
1975b Auditory information masking. *Journal of the Acoustical Society of America* 57(S1):S5.
1976a Multidimensional encoding within the temporal microstructure of auditory displays. III. Multistate displays. *Journal of the Acoustical Society of America* 59:148–152.
1976b Identification of random auditory waveforms III. Effect of interference. *Journal of the Acoustical Society of America* 60:680–686.
1978 Periodicity measures for repeated random auditory patterns. *Journal of the Acoustical Society of America* 63:1132–1141.
1983 Temporal microstructural rules for periodicity of random-amplitude sequences. *Journal of the Acoustical Society of America* 74:S70.
Pollack, I., and L. Ficks
1954 Information of elementary multidimensional displays. *Journal of the Acoustical Society of America* 26:155–158.
Pollard, H.F., and E.V. Jansson
1982b A tristimulus method for specification of musical timbre. *Acustica* 51:162–171.
Pols, L.C.W., L.J. van der Kamp, and R. Plomp
1969 Perceptual and physical space of vowel sounds. *Journal of the Acoustical Society of America* 46(2):458–467.
Pomerantz, J.R., and M. Kubovy
1981 Perceptual organization: An overview. In M. Kubovy and J.R. Pomerantz (Eds.), *Perceptual Organization.* Hillsdale, N.J.: Lawrence Erlbaum.
1986 Theoretical approaches to perceptual organization. In K.R. Boff, L. Kaufman,and J.P. Thomas (Eds.), *Handbook of Human Perception and Performance.* New York: John Wiley.
Powers, G.L., and J.C. Wilcox
1977 Intelligibility of temporally interrupted speech with and without intervening noise. *Journal of the Acoustical Society of America* 61:195–199.
Preis, A.
1980 The method of complex tone classification according to the envelope shape of their spectra. *Acustica* 3(7):136–140.
1984 An attempt to describe the parameter determining the timbre of steady-state harmonic complex tones. *Acustica* 55(1):1–13.
Preusser, D.
1972 The effect of structure and rate on the recognition and description of auditory temporal patterns. *Perception and Psychophysics* 11:233–240.
Purks, S.R., D.J. Callahan, L.D. Braida, and N.I. Durlach
1980 Intensity perception X: Effect of preceding stimulus on identification performance. *Journal of the Acoustical Society of America* 67:634–637.
Pynn, C.T., L.D. Braida, and N.I. Durlach
1972 Intensity perception III: Resolution in small range identification. *Journal of the Acoustical Society of America* 51:559–566.

Raab, D.H., and I.A. Goldberg

1975 Auditory intensity discrimination with bursts of reproducible noise. *Journal of the Acoustical Society of America* 57:437–447.

Rabiner, L., and S. Levinson

1981 Isolated and connected word recognition: Theory and selected applications. *IEEE Trans.* ASSP, ASSP-29:254–272.

Rabinowitz, W.M.

1970 Frequency and Intensity Resolution in Audition. S.M. Thesis, Department of Electrical Engineering, Massachusetts Institute of Technology.

Rabinowitz, W.M., J.S. Lim, L.D. Braida, and N.I. Durlach

1976 Intensity perception VI. Summary of recent data on deviations from Weber's law for 1000-Hz tone pulses. *Journal of the Acoustical Society of America* 59:1506–1509.

Raphael, L.

1972 Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America* 51:1296–1303.

Rasch, R.A.

1978 The perception of simultaneous notes as in polyphonic music. *Acustica* 40:21–33.

1979 Synchronization in performed ensemble music. *Acustica* 43:121–131.

Rasch, R.A., and R. Plomp

1982 The listener and the acoustic environment. In D. Deutsch (Ed.), *The Psychology of Music.* New York: Academic Press.

Remez, R.E.

1987a Neural models of speech perception. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition.* New York: Cambridge University Press.

1987b Units of organization and analysis in the perception of speech. Pp. 419–433 in M.E.H. Schouten (Ed.), *The Psychophysics of Speech Perception.* Boston: Martinus-Nijhoff.

Remez, R.E., P.E. Rubin, D.B. Pisoni, and T.D. Carrell

1981 Speech perception without traditional speech cues. *Science* 212:947–950.

Repp, B.H.

1982 Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin* 92:81–110.

1984 Categorical perception: Issues, methods, and findings. In N.J. Lass (Ed.), *Speech and Language: Advances in Basic Research and Practice.* New York: Academic Press.

1987 The sound of two hands clapping: An exploratory study. *Journal of the Acoustical Society of America* 81:1100–1110.

Repp, B.H. and A.M. Liberman

1987 Phonetic category boundaries are flexible. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition.* New York: Cambridge University Press.

Riesz, R.R.

1928 Differential intensity sensitivity of the ear for pure tones. *Physiological Review* 31:867.

Rock, I.

1983 *The Logic of Perception.* Cambridge, Mass.: MIT Press.

Ronken, D.A.

1969 Intensity discrimination of Rayleigh noise. *Journal of the Acoustical Society of America* 45:54–57.

Rosen, S., and P. Howell

1981 Plucks and bows are not categorically perceived. *Perception and Psychophysics* 30:156–168.

1987a Auditory, articulatory, and learning explanations of categorical perception in speech. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition.* New York: Cambridge University Press.

1987b Is there a natural sensitivity at 20 ms in relative tone-onset-time continua? A reanalysis of Hirsh's (1959) data. Pp. 199–209 in M.E.H. Schouten (Ed.), *The Psychophysics of Speech Perception.* Boston: Martinus-Nijhoff.

Rosenblum, L.D., C. Carello, and R.E. Pastore

1987 Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception and Psychophysics* (under review).

Rothkopf, E.Z.

1957 A measure of stimulus similarity and errors in some paired associate-learning tasks. *Experimental Psychology* 53:94–101.

Royer, F.L., and W.R. Garner

1966 Response uncertainty and perceptual difficulty of auditory temporal patterns. *Perception and Psychophysics* 1:41–47.

1970 Perceptual organization of nine-element auditory temporal patterns. *Perception and Psychophysics* 7:115–120.

Samuel, A.G.

1977 The effect of discrimination training on speech perception: Non-categorical perception. *Perception and Psychophysics* 22:321.330.

1981a Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General* 110:474–494.

1981b The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance* 7:1124–2231.

1987 Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Language* 26:36–56.

Samuel, A.G., and W.H. Ressler

1986 Attention within word perception: Insights from the phonetic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance* 12:70–79.

Sasaki, T.

1980 Sound restoration and temporal localization of noise in speech and music sounds. *Tohoku Psychologica Folia* 39:79–88.

Sawusch, J.R., and P. Jusczyk

1981 Adaptation and contrast in the perception of voicing. *Journal of Experimental Psychology: Human Perception and Performance* 7:408–421.

Sawusch, J.R. and J.W. Mullennix

1985 When selective adaptation and contrast effects are distinct: A reply to Diehl, Kluender, and Parker. *Journal of Experimental Psychology: Human Perception and Performance* 11:243–250.

Scharf, B., S. Quigley, C. Aoki, N. Peachey, and A. Reeves

1987 Focused auditory attention and frequency selectivity. *Perception and Psychophysics.*

Scheffers, M.T.M.

1983 Simulation of auditory analysis of pitch: An elaboration on the DWS pitch meter. *Journal of the Acoustical Society of America* 74:1716–1725.

Schiffman, S.S., M.L. Reynolds, and F.W. Young

1981 *Introduction to Multidimensional Scaling: Theory, Methods, and Applications.* New York: Academic Press.

Schmidt-Neilsen, A., and K. Stern

1985 Identification of known voices as a function of familiarity and narrow band coding. *Journal of the Acoustical Society of America* 77:658–663.

Schoeffler, M.S.

1965 Theory for psychophysical learning. *Journal of the Acoustical Society of America* 37:1124–1133.

Schorer, E.

1986 Critical modulation frequency based on detection of AM vs. FM tones. *Journal of the Acoustical Society of America* 79:1054–1057.

Schouten, M.E.H.

1980 The case against a speech mode of perception. *Acta Psychologica* 44:71–98.

1987 *The Psychophysics of Speech Perception.* Boston: Martinus-Nijhoff.

Schubert, E.D., and R.A. West

1969 Recognition of repeated patterns: A study of short term auditory storage. *Journal of the Acoustical Society of America* 46:1493–1501.

Schwab, E.C., H.C. Nusbaum, and D.B. Pisoni

1985 Some effects of training on the perception of synthetic speech. *Human Factors* 27:395–408.

Seashore, C.E.

1932 *The Vibrato.* University of Iowa Studies in the Psychology of Music 1. Iowa City: The University Press.

1935 *The Psychology of Vibrato.* University of Iowa Studies in the Psychology of Music 3. Iowa City: The University Press.

Shannon, R.V., and T. Houtgast

1986 Growth of pulsation threshold of a suppressed tone as a function of its level. *Hearing Research* 21:251–255.

Shepard, R.N.

1962 The analysis of proximities as a technique for the study of information. *Human Factors* 5:33–48.

1972 Psychological representation of speech sounds. Pp. 67–113 in E.E. David and P.B. Denes (Eds.), *Human Communication: A Unified View.* New York: McGraw-Hill.

1982 Structural representations of musical pitch. In D. Deutsch (Ed.), *The Psychology of Music.* New York: Academic Press.

Shipley, E.F.
1965 Detection and recognition: Experiments and choice models. *Journal of Mathematical Psychology* 2:277–311.
Shower, E.G., and R. Biddulph
1931 Differential pitch sensitivity of the ear. *Journal of the Acoustical Society of America* 3:275–277
Siegel, R.A.
1979 Internal and External Noise in Auditory Detection. Masters Thesis, Massachusetts Institute of Technology.
Silverman, E.B., and J.H. Howard
1977 Recognition of Three Distinctive Features in Brief-Duration Complex Non-Speech Sounds. ONR Contract Number N00014-75-C-0308.: ONR-77-3. September.
Sneath, P.H.A., and R.R. Sokal
1973 *Numerical Taxonomy: The Principles and Practice of Numerical Classification.* San Francisco: Freeman.
Soderquist, D.R.
1970 Frequency analysis and the critical band. *Perception and Psychophysics* 21:117–119.
Sokal, R.R., and P.H.A. Sneath
1963 *Principles of Numerical Taxonomy.* San Francisco: Freeman.
Soli, S.D., P. Arabie, and J.D. Carroll
1986 Discrete representation of perceptual structure underlying consonant confusions. *Journal of the Acoustical Society of America* 79:826–837.
Sorkin, R.D.
1962 Extension of the theory of signal detectability to matching procedures in psychoacoustics. *Journal of the Acoustical Society of America* 34:1745–1751.
1987 Temporal factors in the discrimination of tonal sequences. *Journal of the Acoustical Society of America* 82:1218–1226.
Sorkin, R.D., and M. Snow
1987 Discrimination of time-expanded tonal sequences. *Jounal of the Acoustical Society of America* 82:S40.
Sorkin, R.D., R.E. Pastore, and J.D. Gilliom
1968 Signal probability and the listening band. *Perception and Psychophysics* 4:10–12.
Spiegel, M.F., and D.M. Green
1981 Two procedures for estimating internal noise. *Journal of the Acoustical Society of America* 70:69–73.
1982 Signal and masker uncertainty with noise maskers of varying duration, bandwidth, and center frequency. *Journal of the Acoustical Society of America* 71:1204–1210.
Spiegel, M.F., M.C. Picardi, and D.M. Green
1981 Signal and masker uncertainty in intensity discrimination. *Journal of the Acoustical Society of America* 70:1015–1019.
Spiegel, M.F., and C.S. Watson
1981 Factors in the discrimination of tonal patterns. III. Frequency discrimination with components of well learned patterns. *Journal of the Acoustical Society of America* 69:223–230.
1984 Performance on frequency-discrimination tasks by musicians and nonmusicians. *Journal of the Acoustical Society of America* 76:1690–1695.
Stevens, K.N.
1980 Acoustic correlates of some phonetic categories. *Journal of the Acoustical Society of America* 68:836–842.
Stevens, K.N., and S.E. Blumstein
1978 Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America* 64:127–140.
1981 The search for invariant acoustic correlates of phonetic features. In P.D. Emias and J.L. Miller (Eds.), *Perspectives on the Study of Speech.* Hillsdale, N.J.: Lawrence Erlbaum.
Stevens, S.S., and E.B. Newman
1936 The localization of actual sources of sound. *American Journal of Psychology* 48:297–306.
Strange, W.
1972 The Effects of Training on the Perception of Synthetic Speech Sounds: Voice Onset Time. Ph.D. Dissertation, University of Minnesota.
1986 Speech input and the development of speech perception. In *Otitis Media and Child Development.* Parkton, MJ.: York Publishers.

Strange, W., and S. Dittmann
1984 Effects of discrimination training of the perception of /r-l/ by Japanese Adults learning English. *Perception and Psychophysics* 36:131–145.

Strange, W., and J.J. Jenkins
1978 Role of linguistic experience in the perception of speech. In R.E. Walk and H.L. Pick (Eds.), *Perception and Experience*. New York: Plenum Press.

Studdert-Kennedy, M., A.M. Liberman, K.S. Harris, and F.S. Cooper
1970 The motor theory of speech perception: A reply to Lane's critical review. *Psychological Review* 77:234–249.

Summerfield, Q.
1981 Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 7:1074–1095.
1982 Differences between spectral dependencies in auditory and phonetic temporal processing: Relevance to the perception of voicing in initial stops. *Journal of the Acoustical Society of America* 72:51–61.

Summerfield, Q., A. Sidwell, and T. Nelson
1987 Auditory enhancement of changes in spectral amplitude. *Journal of the Acoustical Society of America* 81:700–708.

Swets, J.A.
1984 Mathematical models of attention. Pp. 183–242 in R. Parasuraman and D.R. Davies (Eds.), *Varieties of Attention*. New York: Academic Press.

Swets, J.A., D.M. Green, D.J. Getty, and J.B. Swets
1978 Signal detection and identification at successive stages of observation. *Perception and Psychophysics* 24:275–289.

Swets, J.A., S.H. Millman, W.E. Fletcher, and D.M. Green
1962 Learning to identify nonverbal sounds: An application of a computer as a teaching machine. *Journal of the Acoustical Society of America* 34:928–935.

Swets, J.A., E.F. Shipley, M.J. McKey, and D.M. Green
1959 Multiple observations of signals in noise. *Journal of the Acoustical Society of America* 31:514–521.

Tallal, P., and M. Piercy
1974 Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia* 12:83–93

Tanner, W.P.
1956 Theory of recognition. *Journal of the Acoustical Society of America*, 28:882–888.
1961 Physiological implications of psychophysical data. *Annals of the New York Academy of Sciences* 89:752–765.

Tanner, W.P., and R.Z. Norman
1954 The human use of information II. Signal detection for the case of an unknown signal parameter. *Transactions of the Institute of Radio Engineers, Professional Group on Information Theory*, PGIT-4:222–227.

Taylor, M.M., and C.D. Creelman
1967 PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America* 41:782–787.

Tees, R.C., and J.F. Werker
1984 Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology* 38:579–590.

Teranishi, R.
1977 Critical rate for identification and information capacity in hearing system. *Journal of the Acoustical Society of Japan* 33:136–143.

Terhardt, E., G. Stoll, and M. Sweewann
1982 Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America* 71:679–688.

Thomas, I.B., and P.J. Fitzgibbons
1971 Temporal order and perceptual classes. *Journal of the Acoustical Society of America* 50:86–87 (abstract).

Thomas, I.B., R.P. Cetti, and P.W. Chase
1971 Effect of silent intervals on the perception of temporal order for vowels. *Journal of the Acoustical Society of America* 49:S84 (abstract).

Thomas, I.B., P.B. Hill, F.S. Carroll, and B. Garcia
1970 Temporal order in the perception of vowels. *Journal of the Acoustical Society of America* 48:1010–1013.

Thurlow, W.R.
1957 An auditory figure-ground effect. *American Journal of Psychology* 70:653-654.
Thurlow, W.R., and L.F. Elfner
1959 Continuity effects with alternately sounding tones. *Journal of the Acoustical Society of America* 31:1337–1339.
Thurlow, W.R. and W.P. Erchul
1978 Understanding continuity effects with complex stimuli. *Journal of the American Auditory Society* 4:113–116.
Thurlow, W.R. and A.E. Marten
1962 Perception of steady and intermittent sound with alternating noise-burst stimuli. *Journal of the Acoustical Society of America* 34:1853–1858.
Treisman, M.
1984 A theory of criterion setting: An alternative to the attention band and response ratio hypothesis in magnitude estimation and cross-modality matching. *Journal of Experimental Psychology: General* 113:443–463.
1985 The magical number seven and some other features of absolute judgment: An application of criterion-setting theory. *Journal of Mathematical Psychology* 29:175–230.
Treisman, M., and A. Faulkner
1984 The setting and maintenance of criteria representing levels of confidence. *Journal of Experimental Psychology: Human Perception and Performance* 10:119–139.
1985 Can decision criteria interchange locations? Some positive evidence. *Journal of Experimental Psychology: Human Perception and Performance* 11:187–208.
Treisman, M., and T.C.A. Williams
1984 A theory of criterion setting. I. Sequential dependencies. *Psychological Review* 91:68–111.
VanDerveer, N.J.
1979 Acoustic Information for Event Perception. Paper presented at the celebration in honor of Eleanor J. Gibson, Cornell University, Ithaca, N.Y.
van Noorden, L.P.
1975 Temporal Coherence in the Perception of Tone Sequences. Thesis, University of Amsterdam.
van Ryzin, J.
1977 *Classification of Clustering*. New York: Academic Press.
von Bekesy, G.
1963 Three experiments concerned with pitch perception. *Journal of the Acoustical Society of America* 35:602–606.
von Bismarck, G.
1974a Timbre of steady sounds: A factorial investigation of its verbal attributions. *Acustica* 30:146–159.
von Bismarck, G.
1974b Sharpness as an attribute of the timbre of steady sounds. *Acustica* 30:159–172.
Verschuure, J.
1978 Auditory Excitation Patterns: The Significance of the Pulsation Threshold Method for the Measurement of Auditory Nonlinearity. Ph.D. Dissertation, Erasmus University, Rotterdam, Netherlands.
Verschuure, J., M. Rodenburg, and A.J.J. Maas
1974 Frequency selectivity and temporal effects of the pulsation threshold method. *Proceedings of the 8th International Congress on Acoustics* (London) 1:131.
Vicario, C.
196 L'effetto Tunnel Acustico. *Rivista di Psicologia* 54:51–52.
Wakefield, G., and N.F. Viemeister
1985 Temporal interaction between pure tone and modulated noise. *Journal of the Acoustical Society of America* 77(4):1535–1542.
Walley, A.C., D.B. Pisoni, and R.N. Aslin
1981 The role of early experience in the development of speech perception. In R.N. Aslin, J.R. Alberts,and M.R. Petersen (Eds.), *Development of Perception: Psychobiological Perspectives Audition, Somatic Perception, and the Chemical Senses*. New York: Academic Press.
Warren, R.M.
1968 Relation of verbal transformations to other perceptual phenomena. *Conference Publication No. 42, Institution of Electrical Engineers* (London): Supplement No. 1:1–8.
1970 Perceptual restoration of missing speech sounds. *Science* 167:392–393.
1972 Perception of temporal order: Special rules for initial and terminal sounds of sequences. *Journal of the Acoustical Society of America* 51:167 (abstract).

1974 Auditory temporal discrimination by trained listeners. *Cognitive Psychology* 6:273–256.
1982 *Auditory Perception: A New Synthesis.* New York: Pergamon Press.
1984 Perceptual restoration of obliterated sounds. *Psychological Bulletin* 96:371–383.
1985 Criterion shift rule and perceptual homeostasis. *Psychological Review* 92:574–584.
1986 Acoustic sequences: Explaining a perceptual paradox. *Journal of the Audio Engineering Society* 34:1021.

Warren, R.M., and J.M. Ackroff
1976 Two types of auditory sequence perception. *Perception and Psychophysics* 20:387–394.

Warren, R.M., and J.A. Bashford
1981 Perception of acoustic utterance: Pitch and infrapitch. *Perception and Psychophysics* 29:395–402.

Warren, R.M., and D.L. Byrnes
1975 Temporal discrimination of recycled tonal sequences: Pattern matching and naming of order by untrained listeners. *Perception and Psychophysics* 18:273–280.

Warren, R.M., and M.D. Meyers
1987 Effects of listening to repeated syllables: Category boundary shifts versus verbal transformations. *Journal of Phonetics* 15:169–181.

Warren, R.M., and C.J. Obusek
1971 Speech perception and phonemic restorations. *Perception and Psychophysics* 9:358–362.

Warren, W.H., and R.R. Verbrugge
1984 Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology : Human Perception and Performance* 10:704–712.

Warren, R.M., and R.P. Warren
1970 Auditory illusions and confusions. *Scientific American* 223:30–36.

Warren, R.M., C.J. Obusek, and J.M. Ackroff
1972 Auditory induction: Perceptual synthesis of absent sounds. *Science* 176:1149–1151.

Warren, R.M., C.J. Obusek, R.M. Farmer, and R.P. Warren
1969 Auditory sequence: Confusions of patterns other than speech or music. *Science* 164:586–587.

Warren, W.H., E.E. Kim, and R. Husney
1987 The way the ball bounces: Visual and auditory perception of elasticity and control of the bounce pass. *Perception.*

Watson, C.S.
1976 Factors in the discrimination of word-length auditory patterns. Pp. 175–188 in S.K. Hirsh, D.H. Eldredge, and S.R. Silverman (Eds.), *Hearing and Davis: Essays Honoring Hallowell Davis.* St. Louis, Mo.: Washington University Press.
1980 Time course of auditory perceptual learning. *Annals of Otology, Rhinology and Laryngology* 74:96–102.
1987 Uncertainty, informational masking, and the capacity of immediate recall. Pp. 267–277 in W.A. Yost and C.S. Watson (Eds.), *Auditory Processing of Complex Sounds.* Hillsdale, N.J.: Lawrence Erlbaum.

Watson, C.S., and D.C. Foyle
1985 Central factors in the discrimination and identification of complex sounds. *Journal of the Acoustical Society of America* 78(1):375–380.

Watson, C.S., and W.J. Kelly
1978 Informational masking in auditory patterns. *Journal of the Acoustical Society of America* 64:S39.
1981 The role of stimulus uncertainty in the discrimination of auditory patterns. Pp. 37–59 in D.J. Getty and J.H. Howard (Eds.), *Auditory and Visual Pattern Recognition.* Hillsdale, N.J.: Lawrence Erlbaum.

Watson, C.S., W.J. Kelly, and H.W. Wroton
1976 Factors in the discrimination of tonal patterns. II Selective attention and learning under various levels of stimulus uncertainty. *Journal of the Acoustical Society of America* 60:1176–1186.

Watson, C.S., H.W. Wroton, W.J. Kelly, and C.A. Benbasset
1975 Factors in the discrimination of tonal patterns. I. Component frequency, temporal position, and silent intervals. *Journal of the Acoustical Society of America* 57:1175–1185.

Weber, D.L.
1983 Do off-frequency simultaneous maskers suppress the signal? *Journal of the Acoustical Society of America* 73:887–893.

Weber, D.L., D.M. Green, and R.D. Luce
1977 Effects of practice and distribution of auditory signals on absolute identification. *Perception and Psychophysics* 22:223–231.

Webster, J.C.
1981 Meaningfulness and the perception of complex sounds. Pp. 61–78 in D.J. Getty and J.H. Howard (Eds.), *Auditory and Visual Pattern Recognition.* Hillsdale, N.J.: Lawrence Erlbaum.

Webster, J.C., A. Carpenter, and M.M. Woodhead
1968a Identifying meaningless tonal complexes. *Journal of the Acoustical Society of America* 44:606–609.
1968b Identifying meaningless tonal complexes II. *Journal of Auditory Research* 8:251–260.

Webster, J.C., M.M. Woodhead, and A. Carpenter
1973 Perceptual confusions between four-dimensional sounds. *Journal of the Acoustical Society of America* 53(2): 448–456.

Wedin, L., and G. Goude
1972 Dimension analysis of the perception of instrumental timbre. *Scandinavian Journal of Psychology* 13:228–240.

Werker, J.F., and J.S. Logan
1985 Cross-language evidence for three factors in speech perception. *Perception and Psychophysics* 37:35–44.

Werker, J.F., and R.C. Tees
1984 Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology* 37:1866–1878.

Wessel, D.L.
1979 Timbre space as a musical control structure. *Computer Music Journal* 3(2):45–52.

Wier, C.C. and D.M. Green
1975 Temporal acuity as a function of frequency difference. *Journal of the Acoustical Society of America* 57:1512–1515.

Wightman, F.L.
1973 The pattern-transformation model of pitch. *Journal of the Acoustical Society of America* 54(2):407–416.

Wightman, F.L., D. Kistler, and M. Perkins
1987 New technique for studies of localization. In W.A. Yost and G. Gourevitch (Eds.), *Directional Hearing.* New York: Springer-Verlag.

Wiley, R.L.
1968 Speech Communication Using the Strongly Voiced Components Only. Ph.D. Dissertation, Imperial College, University of London.

Williams, L.
1979 The modification of speech perception and production in second-language learning. *Perception and Psychophysics* 26:95–104.

Winkel, F.
1967 *Music, Sound, and Sensation: A Modern Exposition.* New York: Dover.

Wish, M.
1967 A model for the perception of Morse code-like signals. *Human Factors* 9:529–540.

Wood, C.C.
1975 Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analysis. *Journal of Experimental Psychology: Human Perception and Performance* 1:3–20.
1976 Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America* 60:1381–1389.

Yost, W.A.
1982 The dominance region and ripple noise pitch: A test of peripheral weighting model. *Journal of the Acoustical Society of America* 72(2):416–425.

Yost, W.A., and G. Gourevitch
1986 *Directional Hearing.* New York: Springer-Verlag.

Yost, W.A., and E.R. Hafter
1987 Lateralization. In W.A. Yost and G. Gourevitch (Eds.), *Directional Hearing.* New York: Springer-Verlag.

Yost, W.A., and M.J. Moore
1987 Temporal changes in a complex spectral profile. *Journal of the Acoustical Society of America* 81:1896–1905.

Yost, W.A., and D.W. Nielsen
1985 *Fundamentals of Hearing: An Introduction.* New York: Holt Rinehart & Winston.

Yost, W.A., and S. Sheft
1988 Modulation detetection interference. *Journal of the Acoustical Society of America.* Under review.

AD-A209 141 CLASSIFICATION OF COMPLEX NONSPEECH SOUNDS PANEL ON CLASSIFICATION OF COMPLEX NONSPEECH SOUNDS(U) NATIONAL RESEARCH COUNCIL WASHINGTON DC W YOST ET AL. 14 APR 89 2/2

UNCLASSIFIED N00014-87-C-0247 F/G 20/1 NL

Yost, W.A., and C.S. Watson

1987a An introduction. In W.A. Yost and C.S. Watson (Eds.), *Auditory Processing of Complex Sounds*. Hillsdale, N.J.: Lawrence Erlbaum.

1987b *Auditory Processing of Complex Sounds*. Hillsdale, N.J.: Lawrence Erlbaum.

Yost, W.A., P.J. Harder, and R. H. Dye

1987 Complex spectral patterns with interaural differences: Dichotic pitch and the 'central spectrum.' In W.A. Yost and C.S. Watson (Eds.), *Auditory Processing of Complex Sounds* Hillsdale N.J.: Lawerence Erlbaum.

Yund, E.W., and R. Efron

1974 Dichoptic and dichotic micropattern discrimination. *Perception and Psychophysics* 15:383–390.

Zurek, P.

1987 The precedence effect. In W.A. Yost and G. Gourevitch (Eds.), *Directional Hearing*. New York: Springer-Verlag.

Zwicker, E.

1952 Die Grenzen der Horbarkeit der Amplitudenmodulation und der Frequenzmodulation Eines Tones. *Acustica* 2:125133.

DISTRIBUTION LIST

Dr. Earl Alluisi
Office of the Deputy Under Secretary of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D.C. 20301

Naval Aerospace Medical Research Lab
Sensory Division Code 23
Pensacola, FL 32508

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, MD 21402

Director
Technical Information Division
Code 2627
Naval Research Laboratory
Washington, D.C. 20375-5000

Dr. Richard Doolittle
Code 5120
Naval Research Laboratory
Washington, D.C. 20375-5000

Mr. Jeff Grossman
Naval Ocean System Center
Code 4403 Bldg 334
San Diego, CA 92152

CAPT W. Moroney
Naval Air Development Center
Code 602
Warminster, PA 18974

Dr. H.C. Okraski
Director, R&D Dept.
Naval Training Systems Center
Orlando, FL 32813-7000

Perceptual Science Program
Office of Naval Research
Code 1142PS
800 North Quincy Street
Arlington, VA 22217-5000 (3 copies)

Dr. Jerry Tobias
NSMRL
Auditory & Communication Sciences Dept.
Naval Submarine Base
Box 900
Groton, CT 06349-5900

Special Assistant for Marine Corps Matters
Code OOMC
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

Dr. Michael Drillings
Basic Research Office
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. Kenneth R. Boff
AF AMRL/HE
Wright-Patterson AFB, OH 45433

Dr. J. Tangney
Life Sciences Directorate, AFSOR
Bolling AFB, Bldg. 410
Washington, D.C. 20032-6448

Defense Technical Information Center
Cameron Station, Bldg. 5
Alexandria, VA 22314 (2 copies)

Dr. Richard Loutitt
Division of Behavioral and Neural Sciences
National Science Foundation
1800 G. Street, N.W.
Washington, D.C. 20550

Dr. James Ballas
Department of Psychology
George Mason University
4400 University Drive
Fairfax, VA 22030

Dr. James H. Howard, Jr.
Department of Psychology
Catholic University
Washington, D.C. 20064

Dr. James A. Simmons
Department of Psychology
Brown University
Providence, R.I. 02912

Dr. Wallace Sinaiko
Smithsonian Institution
Manpower R&D Program
801 N. Pitt Street, Suite 120
Alexandria, VA 22314-1713

Dr. H. P. Van Cott
NAS-National Research Council
(COHF)
2101 Constitution Avenue, N.W.
Washington, D.C. 20418